

AD-A237 005



The Roles of Similarity in Transfer:  
Determinants of Similarity-based  
Reminding and Mapping

Dedre Gentner  
University of Illinois

Mary Jo Rattermann  
University of Illinois

Cognitive Science  
Technical Report CS-91-12  
(Learning Series)

**The Roles of Similarity in Transfer:  
Determinants of Similarity-based  
Reminding and Mapping**

Dedre Gentner  
University of Illinois

Mary Jo Rattermann  
University of Illinois

Cognitive Science  
Technical Report CS-91-12  
(Learning Series)

The Beckman Institute  
University of Illinois  
405 North Mathews Avenue  
Urbana IL 61801  
Tel: (217) 244-5582  
Fax: (217) 244-8371

E-mail [jweaver@director.beckman.uiuc.edu](mailto:jweaver@director.beckman.uiuc.edu)

Acquisition For	
Full Text	<input checked="" type="checkbox"/>
Doc. Fax	<input type="checkbox"/>
Microfilm	<input type="checkbox"/>
Justification	
By	
Distribution/	
Availability Codes	
Avail and/or	
Dist	Special
A-1	



**91-02408**



6 17 098

This research was supported by the Office of Naval Research Cognitive Science Program under Contract NO. N00014-89-J-1272, Contract Authority Identification Number, NR 667-551. We thank Ken Forbus, Art Markham, Janice Skorstad, Bob Schumacher, and Cathy Clement for their helpful comments; and Jennifer Glenn and Linda May for their help in preparing the manuscript. Send correspondence to Dr. Dedre Gentner, Department of Psychology, Northwestern University, 2029 Sheridan Road, Evanston, IL 60208.

Reproduction in whole or part is permitted for any purpose of the United States Government.

Approved for public release; distribution unlimited.

Copyright ©1991 by Dedre Gentner and Mary Jo Rattermann  
Issued June, 1991

Beckman Institute Technical Reports are printed on recycled and recyclable paper.

## REPORT DOCUMENTATION PAGE

Form Approved  
OMB No. 0704-0188

1a. REPORT SECURITY CLASSIFICATION <b>Unclassified</b>			1b. RESTRICTIVE MARKINGS	
2a. SECURITY CLASSIFICATION AUTHORITY			3. DISTRIBUTION / AVAILABILITY OF REPORT  <b>Approved for public release; distribution unlimited.</b>	
2b. DECLASSIFICATION / DOWNGRADING SCHEDULE				
4. PERFORMING ORGANIZATION REPORT NUMBER(S)  <b>CS-91-12</b>			5. MONITORING ORGANIZATION REPORT NUMBER(S)	
6a. NAME OF PERFORMING ORGANIZATION <b>University of Illinois Beckman Institute</b>		6b. OFFICE SYMBOL (If applicable)	7a. NAME OF MONITORING ORGANIZATION <b>Cognitive Science (Code 1142CS) Office of Naval Research</b>	
6c. ADDRESS (City, State, and ZIP Code) <b>405 N. Mathews Urbana, IL 61801</b>		7b. ADDRESS (City, State, and ZIP Code) <b>800 N. Quincy Street Arlington, VA 22217-5000</b>		
8a. NAME OF FUNDING / SPONSORING ORGANIZATION <b>University of Illinois</b>		8b. OFFICE SYMBOL (If applicable)	9. PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER  <b>N00014-89-J-1272</b>	
8c. ADDRESS (City, State, and ZIP Code) <b>Department of Psychology 603 E. Daniel Champaign, IL 61820</b>		10. SOURCE OF FUNDING NUMBERS		
		PROGRAM ELEMENT NO <b>61153N</b>	PROJECT NO	TASK NO WORK UNIT ACCESSION NO <b>NR442f-007</b>
11. TITLE (Include Security Classification) <b>The Roles of Similarity in Transfer: Determinants of Similarity-based Reminding and Mapping (Unclassified)</b>				
12. PERSONAL AUTHOR(S) <b>Gentner, Dedre and Rattermann, Mary Jo</b>				
13a. TYPE OF REPORT <b>Technical Report</b>		13b. TIME COVERED <b>FROM 88-9-1 TO 92-10-31</b>	14. DATE OF REPORT (Year, Month, Day) <b>91-6-2</b>	15. PAGE COUNT
16. SUPPLEMENTARY NOTATION				
17. COSATI CODES			18. SUBJECT TERMS (Continue on reverse if necessary and identify by block number)	
FIELD	GROUP	SUB-GROUP	<b>Similarity    Analogical mapping    Structure mapping</b> <b>Analogical reminding    Computational simulation</b>	
<b>05</b>	<b>10</b>			
19. ABSTRACT (Continue on reverse if necessary and identify by block number) <p>Similarity is generally agreed to be important in transfer. But recent research suggests that its role(s) are more complex than was originally thought. The present research attempts to isolate and compare the determinants of similarity-based access to memory and the determinants of the subjective soundness of a similarity match. In three studies, subjects first read a large set of stories. Later, they were given new stories that resembled the original stories in different ways. Some of the matches were structural analogs of the first scenarios, while others were surface matches, sharing object attributes but not higher-order relational structure. Subjects were told to write out any of the original stories that came to mind. After completing the reminding task, subjects rated all the story pairs for soundness: i.e., how well inferences true of one story would apply to the other. Structure-mapping theory predicts that soundness will be determined by the degree of structural overlap, including that among higher-order relational structure.</p>				
20. DISTRIBUTION / AVAILABILITY OF ABSTRACT <input checked="" type="checkbox"/> UNCLASSIFIED/UNLIMITED <input type="checkbox"/> SAME AS RPT <input type="checkbox"/> DTIC USERS			21. ABSTRACT SECURITY CLASSIFICATION <b>Unclassified</b>	
22a. NAME OF RESPONSIBLE INDIVIDUAL <b>Dr. Susan Chipman</b>			22b. TELEPHONE (Include Area Code) <b>703-696-4318</b>	22c. OFFICE SYMBOL <b>ONR 1142CS</b>

### Abstract

Similarity is generally agreed to be important in transfer. But recent research suggests that its role(s) are more complex than was originally thought. The present research attempts to isolate and compare the determinants of similarity-based *access* to memory and the determinants of the subjective *soundness* of a similarity match.

In three studies, subjects first read a large set of stories. Later, they were given new stories that resembled the original stories in different ways. Some of the matches were structural analogs of the first scenarios, while others were surface matches, sharing object attributes but not higher-order relational structure. Subjects were told to write out any of the original stories that came to mind. After completing the reminding task, subjects rated all the story pairs for *soundness*: i.e., how well inferences true of one story would apply to the other.

Structure-mapping theory predicts that soundness will be determined by the degree of structural overlap, including that among higher-order relational structure. Consistent with this prediction, across all three studies, structural matches were rated as more sound than surface matches. However, the ordering was reversed in the reminding task: Subjects recalled many more surface matches than analogical matches. The focus on structural overlap in soundness judgments was mirrored by the results of a computer simulation of analogical mapping and evaluation. Running the simulation as a surface matcher, we were also able to mimic the accessibility ordering found among human subjects. Thus, these results point to a dissociation between the kind of similarity that promotes similarity-based access and the kind that promotes analogical mapping and inference. Implications for theories of transfer and case-based reasoning are discussed.

## The Roles of Similarity in Transfer: Determinants of Similarity-based Reminding and Mapping

Analogy and similarity are generally agreed to be central in the transfer of learning (Brown, Bransford, Ferrara, & Campione, 1983; Ellis, 1965; Osgood, 1949). In similarity-based transfer, a person who is considering some current situation is reminded implicitly or explicitly of a prior similar situation, and knowledge about the prior situation is applied to the current situation. For example, people solve problems better if they have solved prior similar problems (Anderson, Farrell, & Sauers, 1984; Pirolli, 1985; Ross, 1987, 1989; Simon & Hayes, 1976). There are also examples of more distant similarity-based transfer. For example, a person told of the surprising death of a seemingly healthy race horse may be reminded of the sudden death of the runner Jim Fixx (Kass, Leake, & Owens, 1987); or a dispute over the disposition of the Sinai might remind someone of an argument about possession of an orange (Kolodner, Simpson, & Sycara-Cyranski, 1985). In cases like these, the similarity that supports transfer resides in common systems of relations: common plans and goals, common causal structures, and so on (e.g., Carbonell, 1981, 1983; Schank, 1982).

The preceding examples demonstrate that transfer can be promoted by different kinds of similarity. However, although the relation between similarity and transfer may be strong, it is not simple. Osgood (1949) argued that transfer can be positive or negative, according to whether the stimuli (in an S-R pair) or the responses are similar from training to test. Ellis (1965), while agreeing that similarity is fundamental to transfer, noted a further complexity, namely, the difficulty in defining and measuring similarity. Recent research has brought home the point concerning the complexity of similarity's role in transfer. In Gick and Holyoak's (1980, 1983) insightful series of studies, subjects were given Duncker's (1945) radiation problem: How can one cure an inoperable tumor when enough radiation to kill the tumor would also kill the surrounding flesh? The solution, which is to converge on the tumor with several weak beams of radiation, is normally discovered by only about 10% of the subjects. But if given a prior analogous story in which soldiers converged on a fort, three times as many subjects (about 30%) produced the correct answer. This indicates that spontaneous analogical transfer can occur to good effect, but it also reveals a limitation, for the great majority of the subjects failed to retrieve the analogous story. Yet, when given a general hint that the story might be relevant, about 75% solved the problem indicating that the failure was not of storage but of access.<sup>1</sup>

There is also evidence that similarity-based reminders can be based on superficial commonalities as well as, or instead of, structural commonalities (Holyoak & Koh, 1987; Novick, 1988; Reed, Ernst, & Banerji, 1974). For example, Ross (1984) taught subjects six probability principles in the context of different story lines (e.g., a drunk and his keys). Subjects were then tested on problems whose story line was similar to that of the appropriate principle, similar to that of an inappropriate principle, or unrelated to any of the study problems. Compared to the unrelated baseline, surface similarity in story lines improved performance when linked to

<sup>1</sup>It has been pointed out (e.g., Holyoak & Thagard, 1989a; Keane, 1985; Wolstencroft, 1989) that this analogy is structurally flawed since the general is trying to protect his troops from being harmed by their surround (the populace) and the surgeon is trying to protect the surround (the flesh) from being harmed by the rays. However, replications that address this issue have produced much the same results.

an appropriate principle and hurt performance when linked to an inappropriate principle. In later research, Ross (1989) found that similarities in story lines had a large influence on the probability of accessing the prior problem, but little influence on the probability of correctly applying the principle.

Such findings lead us to suspect that the 'similarity' in similarity-based access may not be the same as the 'similarity' that supports mapping and inference. We will suggest a double decomposition (Gentner, 1982, 1989), distinguishing (1) multiple *subprocesses* of similarity-based transfer and (2) multiple *classes* of *similarity*. It is generally agreed that similarity-based transfer can be decomposed into a set of (not necessarily temporally ordered) subprocesses (J. Clement, 1986; Gentner, 1988a; Gentner & Landers, 1985; Hall, 1989; Holyoak & Thagard, 1989b; Kedar-Cabelli, 1988; Winston, 1980). Given a current (target) topic in working memory, similarity-based transfer requires (1) *accessing* a prior potential analogy, (2) *mapping* the prior (base) analog onto the target, (3) *evaluating* the soundness of the analog, (4) *computing inferences* in the target (in the case of problem-solving, *applying* the analogy in the target), and, possibly, (5) *extracting the common structure* for later use.

To make distinctions among different similarity classes, we use a decomposition of similarity based on structure-mapping theory (Gentner, 1980, 1983, 1987, 1989). We begin by distinguishing *analogy* from other kinds of similarity. Analogy is defined as similarity in relational structure; more specifically, analogy is a one-to-one mapping from one domain representation (the base) into another (the target) which conveys that a system of relations that holds among the base objects also holds among the target objects independently of any similarities among the objects to which those relations apply.<sup>2</sup> An important assumption is that a further selection is made: among the common relations that could participate in the analogy, people prefer to focus on interconnected systems of relations (the *systematicity principle*). That is, they prefer to map sets of relations which include common higher-order relations that constrain the common lower-order relations. Thus, the structure-mapping account predicts that analogical *mapping* and structural *evaluation* of analogies should depend on the degree to which the analogs share deep relational systems. Given these distinctions, we can distinguish further classes of similarity according to the kinds of predicates that are shared. In *analogy*, only relational predicates – low-order and higher-order – are shared. In *literal similarity* (overall similarity), both relational predicates and object-attributes are shared. In surface, or *mere-appearance* matches, only object-attributes and low-order relations are shared. Table 1 gives examples of these different kinds of similarity.<sup>3</sup>

<sup>2</sup>This represents an ideal competence model of analogy – a description at Marr's (1982) "computational level" or Palmer & Kimchi's (1985) "informational-constraints level." For a discussion of performance factors in analogical processing, see Gentner (1989) or Schumacher & Gentner (1988a, 1988b).

<sup>3</sup>These similarity classes are continua, not dichotomies. Analogy and literal similarity lie on a continuum of degree of attribute overlap. Likewise, there is a continuum between analogy and relational abstraction. Both involve overlap in relational structure, but they vary in the degree of concreteness of the base domain.

### Examples of Different Similarity Types

Table 1

Literal Similarity	Milk is like water
Analogy	Heat is like water
Mere Appearance	The glass tabletop gleamed like water

In order to separate out different kinds of similarity effects, we require an extremely precise methodology. Most studies of similarity-based problem transfer have simply measured the probability of successful transfer (e.g., Gentner & Toupin, 1986; Holyoak & Koh, 1987; Reed, Ernst, & Banerji, 1974; Ross, 1984; Schumacher & Gentner, 1988b). Generally, both kinds of similarity have positive effects on transfer (Novick, 1988). However, overall transfer is far too coarse a measure for our purposes since it involves at least four of the subprocesses discussed above: *accessing*, *mapping*, *evaluating*, and *drawing inferences* from the analogy. One investigation that made finer distinctions is that of Ross (1989), who separated proportion of reminders, as measured by whether subjects wrote out the prior formula (the access process), from proportion of correct solutions. This allows a contrast between solution rate – a measure which presumably includes access to prior problem mapping adaptation, evaluating, and drawing inferences – and reminding rate, which is closer to a pure measure of access. However, since subjects were engaged in trying to solve the current problem, the reminders they wrote out may well have been filtered by the subjects' assessment of their usefulness; thus, the reminding rate may represent the combined effects of mapping and evaluation as well as access.

Another set of studies that attempted to make distinctions in the use of similarity in transfer was performed by Clement and Gentner (1988; 1991). These studies investigated whether the presence or absence of common systematic relational structure would influence people's selection of facts to include in an interpretation of and prediction from an analogy. Subjects were given a pair of analogous scenarios and asked to select which of two common facts belonged best to the analogy. Although both matches were equally close in and of themselves, subjects showed a strong preference for selecting the common fact that belonged to a larger system of common relations. The same preference for connected systems held for subjects' predictions from an analogy. In a second study, subjects were given an analogy and asked to predict one new fact about the target. The materials were designed so that there were two possible predictions (facts represented in the base but not in the target), differing only in whether they were part of a common system of mutually constraining relations. As expected, subjects predicted the fact that was connected to a common system. These results provide evidence for a special role for common relational structure in analogical *mapping*.

There is also evidence concerning the *evaluation* of analogy and metaphor. When asked to rate the aptness of different kinds of metaphors and similes,<sup>4</sup> people rate comparisons based on shared relations (e.g., "A camera is like a tape recorder.") as more apt than those based on

<sup>4</sup>Many of the comparisons were relational metaphors or similes, such as "Sermons are like sleeping pills." (Ortony, 1979), which qualify as analogies.

shared attributes (e.g., "The sun is like an orange."). Further, their aptness ratings are correlated with the degree to which their interpretations include relational (but not attributional) information (Gentner, 1988b; Gentner & Clement, 1988; Gentner, Falkenhainer, & Skorstad, 1988). These studies tell us that there is an overall preference for matching relations over matching object attributes in evaluating metaphoric aptness. However, there are two reasons why they do not provide an adequate test of the systematicity prediction for soundness. First, they utilized *aptness* ratings (explained as "how clever, interesting, or worth reading" the metaphor was) instead of soundness ratings. Aptness may be related to analogical soundness, but it cannot be assumed to be the same thing. (For one thing, aptness may be more sensitive to novelty than soundness.) Second, and more importantly, these results do not address the more specific claim that common *systematic relational structure* is important in subjective goodness: i.e., the claim that analogical soundness is disproportionately increased by the presence of common relational systems composed of lower-order relations and higher-order constraining relations. Therefore the systematicity prediction must be tested using a more direct measure of the subjective soundness of different kinds of matches.

To achieve a more fine-grained account of the role of similarity in different transfer subprocesses, we conducted two separate tasks across the same materials. The tasks were designed to tap into different subprocesses, and hence possibly to reveal different aspects of similarity use. To measure accessibility, we used a story-memory task: Subjects were asked to read a set of stories and later to write out any reminders they experienced when reading a second set of (similar) stories. Their task was thus simply to write out reminders of the first set. This minimizes the contribution of other subprocesses, such as mapping and evaluation, and comes close to being a pure measure of access. To investigate the evaluation process, we presented subjects with the pairs of stories used in the memory task and asked them to rate the soundness (and/or similarity) of the pairs (which should require only mapping and evaluation).

The basic method was the same across all three studies. In order to create relatively realistic long-term memory demands, we first gave subjects a large number of stories to read and remember (the original set). Then, about a week later, we gave them new stories (the cue set) that matched the original stories in various ways and asked them to write out any of the original stories they were reminded of while reading the new stories. In the final step, subjects rated the original and corresponding cue stories for inferential soundness (and/or similarity). The stories making up the original and cue sets were carefully designed to embody different kinds of similarity matches. There were three levels of matching predicates that could be shared by the base (original) and target (cue) stories: (1) matching *object attributes*, leading to object-level similarities (e.g., *countries/nations*); (2) matching *first-order relations*, leading to similarity among events and other relations between objects (e.g., *taking over/overpowering*); and (3) matching *higher-order relations*, leading to similarity in causal structure or in other kinds of plot structure. By combining these three levels of predicate-matching, we constructed various kinds of matches.

In the first experiment,<sup>5</sup> the similarity matches were *surface matches*, which matched in object descriptions and first-order relations, true *analogies*, which matched in first-order and higher-order relations, and *FOR matches*, which matched only in first-order relations. Subjects were told that if the new story reminded them of any of the original stories they should write

---

<sup>5</sup>The first study was carried out by Russell Landers in 1984 as an M.I.T. undergraduate honors project and appears in the Proceedings of the IEEE (Gentner and Landers, 1985).

out that original story in as much detail as possible. After the reminding task, subjects rated the *soundness* of each pair of original and cue stories. Soundness was explained as the degree to which inferences that followed from one story would hold for the other.

Structure-mapping theory predicts that the subjective soundness of an analogy should depend chiefly upon the depth of the common relational structure and not on the amount of overlap in object-attributes. Therefore, it predicts that analogical matches should be rated as more sound than surface matches or FOR matches. The prediction for access is less clear. Some work in case-based reasoning would suggest that higher-order relational commonalities should play an important role in similarity-based access (e.g., Schank, 1982). However, there is also research that suggests that access will be strongly influenced by surface similarity or by both surface and structural similarity (Holyoak & Koh, 1987; Novick, 1988; Ross, 1984, 1987).

## Experiment 1

### Method

#### Subjects

The subjects were 30 students from the MIT Psychology Department.

#### Design and Materials

Design. Similarity type was varied within subjects with each subject receiving 1/3 analogies, 1/3 surface matches, and 1/3 FOR matches. For counterbalancing, subjects were divided into three groups that differed in which stories were presented in each Similarity type. For this purpose the 18 stories were divided into three sets of six. This gave a 3 X 3 Group (between) X Similarity type (within) design. The dependent measures for the reminding task were three measures of subjects' recall of the original stories: judges' ratings of their recalls, proportion of recalls rated above criterion, and proportion of recalls of a predefined key word. The dependent measure for the soundness task was subjects' ratings of the soundness of the matches.

Materials. The materials were 18 sets of stories, with four stories per set, plus 14 filler stories. The stories were two or three paragraphs long. Each of the 18 story sets contained an original story plus three matching cue stories, which differed in the amount and level of similarity they shared with the original (see Table 2). All cues shared identical or nearly identical first-order relations (e.g., events and actions) with the original story, but differed in the other levels of similarity which were shared.

- In analogy (AN) cues, higher-order relations as well as first-order relations matched the original. The objects (i.e., the characters, physical objects, and locations) differed.
- In surface-match (SM) cues, objects and first-order relations matched the original, but the higher-order relational structure (i.e., the causal structure or plot structure) differed.
- In FOR match (FOR) cues, only the first-order relations matched the original. The objects as well as the higher-order structure differed.

**Sample stimuli from Experiment 1****Table 2**Original Story

Two small countries, Bolon and Salam, were adjacent to a large, warlike country called Mayonia. Bolon decided to make the best of the situation by taking over Salam. Salam started looking for aid from other strong countries but soon Bolon succeeded in taking it over. Then victorious Bolon proposed to make a treaty with its warlike neighbor Mayonia. Bolon proposed to give Mayonia control over Salam in exchange for a guarantee that Bolon would remain independent. Mayonia responded by overrunning both Bolon and Salam. Bolon was so busy maintaining control of Salam, it could do nothing to stop Mayonia. Thereupon, Mayonia installed puppet governments in both Bolon and Salam, and took over the newspapers and radio stations.

Analogy Match

Two sixth-grade boys, Lincoln and Moreland, were walking to school together when they met Chad. Chad was a mean high school boy who was known to rob younger kids. Moreland was scared but he decided to make the best of the situation by overpowering Lincoln. Immediately, Lincoln started running to find a policeman but Moreland ran after Lincoln and eventually succeeded in catching him. When Chad arrived Moreland offered to let him rob Lincoln in exchange for being left alone himself. But Chad robbed both of them. Moreland was so busy holding Lincoln down, he could do nothing to prevent it.

Surface Match

Two weak nations, Lincoln and Moreland, bordered a third nation known as Chad. Chad was aggressive and very powerful. Moreland decided to make the best of the situation by offering to help conquer Lincoln in exchange for being left alone itself. But Lincoln somehow received word of the agreement and began to seek protection from other powerful nations. Immediately, Moreland invaded Lincoln and eventually succeeded in conquering it. When Chad got into the action it invaded both Lincoln and Moreland. And Moreland was so drained from battle, it could do nothing to prevent it.

FOR Match

Two seventh graders, Lincoln and Moreland, were walking to school together when they met Chad. Chad was a mean high school boy who was known to rob younger kids. Moreland was scared but he decided to make the best of the situation by offering to help Chad rob Lincoln in exchange for being left alone himself. But Lincoln overheard and started running to find a policeman. Immediately, Moreland took off after him and eventually succeeded in catching him. When Chad arrived he robbed both of them. And Moreland was so exhausted, he could do nothing to prevent it. Then triumphant Moreland offered to make a pact with its aggressive neighbor Chad. Moreland offered to give Chad control over Lincoln in exchange for being left alone itself. When Chad got into the action it invaded both Lincoln and Moreland. And Moreland was so drained from battle, it could do nothing to prevent it.

Each subject received only one matching cue story for each of the 18 original stories. All subjects received the same memory set; the difference was in the type of story used in the cue set. This was done to avoid confounding the issue of interest, namely, whether there is differential reminding according to the kind of similarity between a current item and a prior item with that of whether there is differential *availability* of items in memory.

To ensure comparability, the cue stories in a given set were made as similar as possible, subject to the constraints of the theory. In particular, the AN and FOR cues were constructed to have the same objects, and the SM and FOR cues were constructed to have the same higher-order structure. The objects were people, animals, companies, and countries.

To avoid purely lexical reminders, the use of identical words between original and cue was avoided; similarities were expressed by means of synonyms or closely similar words. (The exceptions were function words and certain conjunctions and prepositions that had no acceptable substitution.) For each story, one to three of the characters were given proper names. These always differed between original and cue but were similar or identical for all three cue stories.<sup>6</sup>

Finally, following a technique used by Read (1983, 1984), we included in each original story a sentence that was not paralleled in any of the cue stories, which served as a pure memory test. This was done because subjects' recall scores for the original stories will be inflated to the extent that they can reconstruct the original purely by transforming the cue. Since the success of such reconstruction techniques could vary across the different match types, it was desirable to have a measure of memory access that could not be mimicked by transforming the cue. This extra sentence was always the final sentence in the original story. It was designed so that (1) it could not be predicted from the rest of the original; (2) it should be interesting enough to remember; (3) it did not alter the basic plot structure of the original; and (4) it contained at least one distinct new word or concept (the key *word* whose presence could be easily detected in the recall). For example, in the 'Bolon and Salam' set shown in Table 1, the final original sentence is "Mayonia installed puppet governments in both Bolon and Salam and took over the newspapers and radio stations" with *puppet governments* as its unique word.

## Procedure

**Reminding task.** Subjects were tested in groups of three to eight in two separate sessions. In the first session, subjects read a booklet containing the 18 original stories intermixed with 14 filler stories. (The first three and last three stories were always fillers.) All subjects read the same 32 stories, in different semi-random orders. They were told to read the stories carefully, so that they would be able to remember them a week later. Subjects took about thirty minutes for this task.

The second session took place six to eight days later. Subjects received a workbook containing 18 cue stories that corresponded to the 18 original stories read in the previous session. Each workbook consisted of six SM cues, six AN cues, and six FOR cues. Subjects were told that for each cue story they should write down any original story of which they were reminded. If they were reminded of more than one story, they were to write the one that best matched the current story. They were told to include as many details as they could remember – if possible,

<sup>6</sup>In some story sets, the sexes of the AN and FOR characters were changed from those of the original to heighten the lower-order differences; the SM characters were, of course, always the same sex as the original. In these cases, the names of the cue characters were kept as similar as possible: e.g., Mr. Boyce and Ms. Boyce, Christian and Christine, and Sidney and Cindy.

the names of characters and their motives as well as the events.

**Soundness-rating task.** Following completion of the reminding task in the second session, the subjects were given a soundness-rating task. In this task they were given the same pairs of stories they had received in the reminding task (regardless of whether the subject had succeeded in the recall task). Thus each subject rated six SM, six AN, and six FOR matches in the same counterbalancing groups as in the reminding task. They were asked to rate each pair for the inferential soundness of the match. In explaining the task, we avoided using terms like "analogy" or "analogous." Instead, *soundness* was explained as follows:<sup>7</sup> "This part of the experiment is about what makes a good match between two stories or situations. We all have intuitions about these things. Some kinds of resemblances seem important, while others seem weak or irrelevant... In this part of the experiment, we want you to use your intuitions about soundness - that is, about when two situations match well enough to make a strong argument... On each page, we will show you a pair of stories. (You may recognize the stories because you have read them before. But don't worry about that.) You will probably notice that the two stories in each pair resemble each other or 'match' in certain ways. We want you to rate how 'sound' the match between the two stories is. A sound match between two stories is one in which the essential aspects of the stories match. To put it another way, a sound match is one in which you can draw conclusions about the second story from the first. If the pair of stories match this way, give them a high rating. The opposite of a sound match is a spurious match. In a spurious match, the resemblance between the stories is shallow. If the pair of stories is like this, give them a low rating..." The subjects were told to rate the pairs of stories on a scale from 1 to 5, with 5 being 'extremely sound' and 1 being 'extremely spurious.'

### Scoring the Reminding Task

The subjects' written recalls were scored by two judges. The judges were provided with a list of the 18 original stories, and for each of the subjects' recalls, the judges were told which original story was the intended match. However, they were not told which cue story the subject had seen; that is, they were blind to the Similarity condition. Three kinds of reminding scores were obtained:

(1) Overall score. The judges rated how close the subject's recalled version was to the actual original story, using a 0-5 scale, as follows:

- 5 = all important elements of the original and many details.
- 4 = all important elements of the original and some details.
- 3 = all important elements of the original but very few or no details.
- 2 = some important elements of the original; others missing or wrong.
- 1 = some elements from the original but not enough to be certain that the subject genuinely recalled the original.
- 0 = no recall or different story.

---

<sup>7</sup>The instructions used here were quite lengthy and are given in full in Appendix A. In subsequent experiments we found that the instructions could be shortened considerably.

The independent scores assigned by the two judges agreed 75.5% of the time and were within one point of each other 97.5% of the time. Disagreements were resolved by discussion.

(2) Proportion reminders (proportion above criterion or flat scoring). The 0-5 recall score reflects the *quality* of the recalls. This could lead to confoundings if some kinds of similarity matches permit more accurate recalls (through reconstruction). Therefore, as a second dependent measure, we dichotomized the judges' scores such that all responses that were judged to be clearly identifiable recalls of the original (i.e., recalls that received an overall score of 2 or better) were classified as reminders while all descriptions with scores of one or zero were classified as non-reminders. This allowed a measure of the number of reminders each subject produced for each of the three Similarity types.

(3) Key word score. As discussed above, as a further means of identifying genuine recalls (as opposed to reconstructions), each original story contained a final key sentence that was not paralleled in the cue stories. Any inclusion of this key sentence was clear evidence of memory for the original. By design, these sentences each contained a distinctive word or concept. Thus, the judges simply noted whether this distinctive key word (or a synonym) occurred in the subject's response. Subjects' descriptions were given a 2 if they contained the key word or a close synonym, a 1 if they contained a dubious synonym, and 0 otherwise.

## Results and Discussion

### Soundness-rating Task

As predicted, subjects judged the pairs that shared higher-order relational structure (the analogy matches) to be significantly more sound than the pairs that did not share higher-order structure (the FOR matches and surface matches). Figure 1a shows the mean soundness ratings for the three types of similarity matches.

This pattern was confirmed by a one-way analysis of variance, which revealed a main effect of Similarity type,  $F(2,58) = 70.51$ ,  $p < .0001$ . Pairwise comparisons using the Bonferroni procedure ( $\alpha = .05$ )<sup>8</sup> to adjust for the multiple comparisons of means showed a significant advantage for AN matches ( $M = 4.40$ ) over SM ( $M = 2.80$ ) and FOR matches ( $M = 2.74$ ) and no significant difference between SM and FOR matches. A one-way ANOVA using items as the random variable revealed a significant effect of Similarity type,  $F(2,34) = 46.306$ ,  $p < .0001$ , indicating that the effect is reliable across items.

### Reminding

The results of the reminding task show a different pattern. The SM matches, which shared object attributes as well as first-order relations, were the most effective reminding cues. The AN matches, which were rated highest in the soundness task, were much less effective than the SM matches in promoting reminding on all three measures. Figure 1b shows the proportion of stories recalled (i.e., the proportion of recalls assigned a score of 2 or better in the judges' ratings) for each type of cue match.

<sup>8</sup> All planned comparisons in this paper utilized the Bonferroni procedure at  $\alpha = .05$  unless otherwise stated.

Mean soundness ratings for the three similarity types in Experiment 1

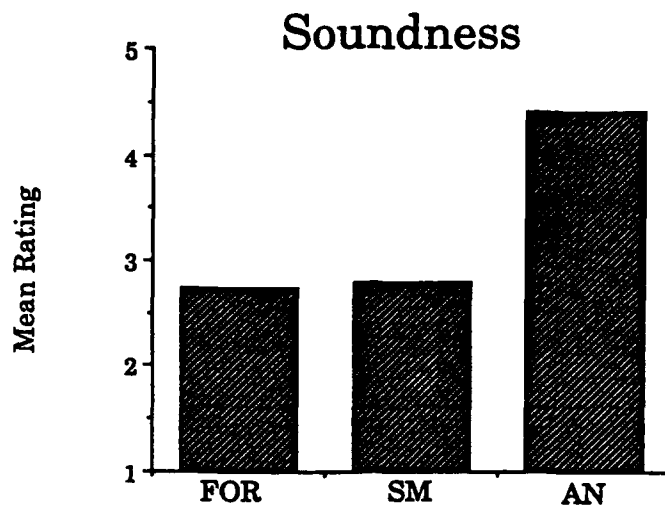


Figure 1a

Proportion recalled for the three similarity types in Experiment 1

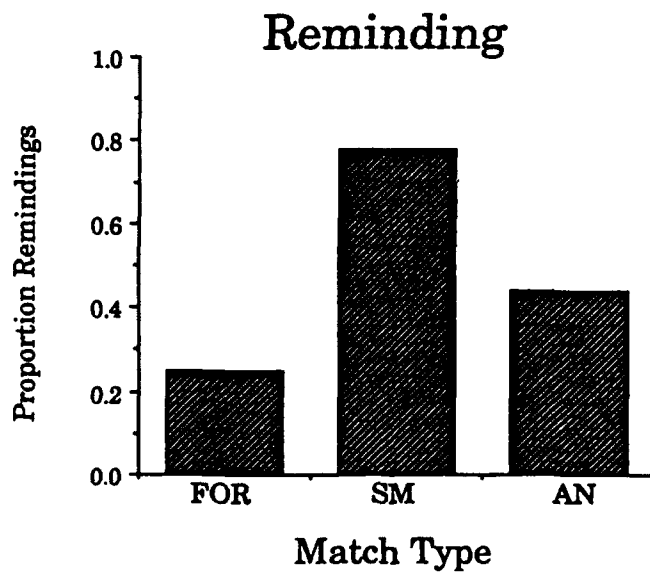


Figure 1b

A one-way analysis of variance confirmed this pattern. There was a significant main effect of Similarity type,  $F(2,58) = 75.45$ ,  $p < .0001$ . Pairwise comparisons ( $\alpha = .05$ ) showed a significant advantage in proportion of reminders for SM matches ( $M = .78$ ) over AN matches ( $M = .44$ ) and for AN matches over FOR matches ( $M = .25$ ). A one-way analysis of Similarity type with items as the random variable showed that the effect of Similarity type is significant across items,  $F(2,34) = 31.80$ ,  $p < .0001$ .

**Results of Experiment 1:**  
**Mean Ratings of soundness and similarity compared to**  
**three measures of reminding across the three similarity types**

Table 3

Measure	Match Type		
	Analogy (R1 + R2) <sup>a</sup>	Surface Matches (OA + R1)	First-order Matches (R1)
Soundness <sup>b</sup>	4.40	2.80	2.74
Proportion Recalled	.44	.78	.25
Proportion of Key Words Recalled	.21	.26	.13
Quality of Reminding	1.48	2.36	.86

Notes:

<sup>a</sup>Here OA refers to object commonalities, R1 to first-order relational commonalities, and R2 to higher-order relational commonalities.

<sup>b</sup>Soundness and quality of reminding are mean ratings on 1 (low) to 5 (high) scales.

The other two measures of reminding showed a similar pattern. Table 3 shows the mean ratings of quality of recall as well as the mean key word scores across the three match types. (For comparison, the proportions of reminders and the mean soundness ratings are also given.)

A one-way analysis of variance performed on the judges' ratings of the recall data revealed the same pattern as in the proportion recalled. There was a main effect of Similarity type,  $F(2,58) = 72.81$ ,  $p < .0001$ . Bonferroni tests showed that the mean rating for SM ( $M = 2.36$ ) exceeded that for AN ( $M = 1.48$ ), which exceeded that for FOR ( $M = .86$ ). A one-way analysis of Similarity type over items also showed a significant effect of Similarity type,  $F(2,34) = 24.79$ ,  $p < .0001$ .

Similar analyses were conducted for the key word score. There was a marginal main effect of Similarity type,  $F(2,58) = 2.80$ ,  $p = .06$ . A one-way analysis of Similarity type over items, however, showed no significant effect of Similarity type,  $F(2,34) = 2.28$ , NS.

Overall, the reminding results present a uniform picture: Surface matches produced the greatest proportion of reminders, followed by analogy and then by FOR matches. It is not surprising that the FOR match is the worst of the three, since it shares fewer commonalities than the other two match types. FOR matches share only first-order relations, while AN matches share higher-order and first-order relations and SM matches share object attributes and first-order relations. Thus, the results can be seen as an indication that (1) adding higher-order relations to a first-order relational match improves reminding (since AN was superior to FOR); (2) similarly, adding common object-attributes to a first-order match improves reminding (since SM was superior to FOR); and more surprisingly, (3) adding common object-attributes to a first-order match improves reminding more than adding common higher-order structure (since SM was superior to AN). These results suggest that common object descriptions may play an important role in similarity-based access.

This pattern differs markedly from the pattern found for subjective soundness. Indeed, the two measures were uncorrelated  $r(52) = .15$ , NS. For subjective soundness, as predicted by structure-mapping theory, common higher-order relational structure is a crucial determinant of the subjective "goodness" of an analogy. Subjects rated AN matches as more sound than FOR matches, indicating that adding higher-order relations to first-order relations improves a match. However, it is not merely that adding features improves soundness, for SM matches (with first-order relations and object attributes) are rated as no better than FOR matches. By this analysis, these results suggest that it is not the mere number of features that determines soundness. Rather, it is the presence of shared *higher-order relational structure* (that is, shared *systematicity*). However, since this conclusion rests on the equivalence of SM and FOR matches (a null result) as well as on the superiority of AN over FOR, it would be prudent to replicate it before drawing strong conclusions.

Perhaps the most surprising aspect of these results is the dissociation between access and inferential soundness: Subjects rated analogical matches as most inferentially sound, but these matches tended not to be retrieved well in the reminding task. Conversely, surface matches were highly retrievable but were not considered sound. This suggests that subjects' memory access mechanisms did not provide them with the matches that they themselves considered most useful in inference. These results suggest that similarity-based access and similarity-based inference may be responsive to different kinds of similarity.

### Experiment 2

Experiment 2 was conducted to replicate and extend Experiment 1. Because of the rather surprising dissociation between access and inferential soundness, a replication is clearly in order. Moreover, an important condition was missing from the similarity design used in Experiment 1: overall (or *literal*) similarity. In Experiment 1, we manipulated relational similarity and object similarity to produce three different match types. Beginning with FOR matches, in which only first-order relations matched, we added either object-attribute commonalities to create surface matches or higher-order relational commonalities to create analogy matches, with the former leading to an access advantage and the latter leading to a soundness advantage. In Experiment 2, we asked what would happen if all three kinds of commonalities – object-attribute commonalities, first-order relational commonalities, and higher-order relational commonalities – were present, creating a literal similarity match. With the addition of Literal Similarity (LS) matches, the set of match types forms a 2 X 2 design, as shown in Figure 2.

Design of materials for Experiment 2

	Common Higher-Order Relations	No Common Higher-Order Relations
Common Object- Attributes	LS	FOR
No Common Object- Attributes	AN	SM

All stories share first-order relations.

Figure 2

This design enables a further test of the predictions for soundness. According to structure-mapping, the subjective soundness of a similarity match depends only on the degree to which relational structure is shared. Therefore, both literal similarity matches and analogies should be rated as highly (and about equally) sound, since both kinds of matches include a substantial common relational structure – both first-order and higher-order relational commonalities. Although structure-mapping predicts that the additional object-attribute commonalities in literal similarity should not contribute to subjective soundness, the literal similarity stories still might receive slightly higher soundness ratings because their first-order events are slightly more similar. In practice, since the choice of objects constrains the choice of allowable relations, the first-order relations were slightly more similar for pairs with similar objects than for pairs with dissimilar objects.

<b>Base</b>	Bolon (a country) <i>takes over</i> Salam (a country).
<b>Literal Sim</b>	Chad (a country) <i>conquers</i> Lincoln (a country).
<b>Analogy</b>	Chad (a boy) <i>robs</i> Lincoln (a boy).

For access, we expected, based on the results of Experiment 1, that literal similarity matches would lead to effective recall in the reminding task, since object commonalities are shared. However, there remains the question of the degree of improvement. There might be essentially additive effects of various kinds and levels of commonality, in which case the addition of higher-order similarity would contribute a constant level of improvement in access. In this case, LS would show the same advantage over SM that AN shows over FOR. A second possibility is that accessibility depends chiefly on lower-order commonalities with higher-order relational similarities contributing little. In this case, LS matches should perform roughly like SM matches in access. Finally, there might be a kind of superadditive effect of combining different levels of similarity, in which case the advantage for LS over SM should be greater than the advantage for MA over FOR.

In the replication, we wished also to address an alternative interpretation of the results of Experiment 1: namely, that the retrievability ordering among the three types of matches was simply a function of their overall similarity. That is, it could have been the case that the surface match cues were simply more similar to their originals than were the analogy cues, and that, in turn, the analogies were more similar to their originals than were the FOR analogies. In this case, there would be no reason to invoke a special role for surface attributes in similarity-based access. To address this problem, we added a similarity-rating task in addition to the soundness-rating task. Finally, to avoid contamination from the cued-recall task to the two ratings tasks and between the two ratings tasks, we conducted the ratings tasks with independent subjects as well as with the recall subjects.

## Experiment 2a: Rating of Soundness and Similarity

### Method

#### Subjects

The subjects were 40 undergraduates at the University of Illinois, 20 of whom received psychology class credit for participation and 20 of whom were paid for their participation. Half the subjects rated soundness and half rated similarity.<sup>9</sup>

#### Materials and Design

The materials were the same as those used in Experiment 1 with two additions. First, two new story sets were added for a total of 20. Second, the story sets were expanded to include, along with the original story, Literal Similarity (LS) cues as well as AN, FOR, and SM cues. This addition required a few minor changes in the prior stories to preserve the parallel design. The LS cues had in common with the original stories all three levels: object attributes, first-order relations, and higher-order relations; SM had two levels (object attributes and first-order relations) as did AN (first-order relations and higher-order relations); and FOR had one level of commonality (first-order relations).

As in Experiment 1, Similarity type was varied within subjects. In order to counterbalance the association of similarity type with particular stories, subjects were divided into four groups each of which received 1/4 of the stories in each of the four similarity conditions. Thus, each story set was rated equally often in each similarity condition, and no subject received more than one pair from any story set. Since the presence of higher-order relations and object attributes varied systematically among the four match types, the design was 2 x 2: Relational commonality (within) x Object commonality (within).

#### Procedure

Soundness-rating task. The instructions for the soundness-rating task were similar to those in Experiment 1.

Similarity-rating task. The method for the similarity-rating task was similar to that used for the soundness ratings: Subjects saw pairs of stories and rated them on a 1-5 point rating scale with 5 = extremely similar and 1 = extremely dissimilar. In the instructions, "similarity" was explained as "overall, how the characters and actions in the two stories resemble one another; or how much there is a general resemblance between the two stories." The vagueness of the description was intentional; we wanted this rating to reflect the subject's own opinion of what makes two items similar.

### Results and Discussion

#### Soundness-rating task

As predicted, subjects based their soundness judgments primarily on the degree to which the two stories shared a relational system. Figure 3a shows the mean soundness ratings for

<sup>9</sup>The 20 paid subjects were evenly distributed between the soundness-rating task and the similarity-rating task.

the four kinds of matches. A 2 x 2 Relational similarity (within) x Object similarity (within) analysis of variance revealed a main effect of Relational similarity,  $F(1,19) = 64.51, p < .0001$ . Neither the main effect of Object similarity nor the interaction of Relational similarity and Object similarity were significant ( $F(1,19) = 2.88, p < .11$  and  $F(1,19) = .006, p < .9$ ). Planned comparisons ( $\alpha = .05$ ) confirmed an advantage for LS matches ( $M = 4.26$ ) and AN matches ( $M = 4.01$ ) over SM matches ( $M = 3.10$ ) and FOR matches ( $M = 2.87$ ). As predicted, there was no significant difference between LS and AN matches nor between MA and FOR matches. An analysis over items also revealed a main effect of Relational similarity,  $F(1,19) = 22.82, p < .0001$ . There was also a main effect of Object similarity,  $F(1,19) = 5.47, p < .05$ . It is possible that this apparent effect of object similarity is in fact due to the slightly better match in first-order relations discussed above. However, overall the pattern of results bears out the prediction that soundness is chiefly determined by the degree of relational match.

### Similarity-rating task

In contrast to their soundness ratings, subjects' similarity ratings were sensitive to both higher-order relational commonalities and object-attribute commonalities. (See Figure 3b.) A similar analysis performed on the similarity ratings revealed main effects of Relational similarity,  $F(1,19) = 85.78, p < .0001$ , and Object similarity,  $F(1,19) = 7.01, p < .02$ . Planned comparisons ( $\alpha = .05$ ) revealed an advantage of LS matches ( $M = 4.50$ ) over AN matches ( $M = 4.09$ ), of AN matches over SM matches ( $M = 3.40$ ), and SM matches over FOR matches ( $M = 2.88$ ). The item analysis also revealed main effects of Relational similarity,  $F(1,19) = 65.25, p < .0001$ , and Object similarity,  $F(1,19) = 5.98, p < .05$ .

The contrast between the patterns for soundness and those for similarity is instructive. Subjects' ratings of soundness were sensitive to relational commonalities but not to object commonalities. Their ratings of literal similarity and analogy matches did not differ, nor did their ratings of surface matches and FOR matches. In contrast, their similarity ratings were sensitive to both object commonalities and relational commonalities. These results suggest that whereas soundness may be a rather pure measure of relational match similarity is multiply determined. Putting it another way, subjective soundness is one contributor to subjective similarity for complex material.

Mean soundness rating for the four similarity types in Experiment 2

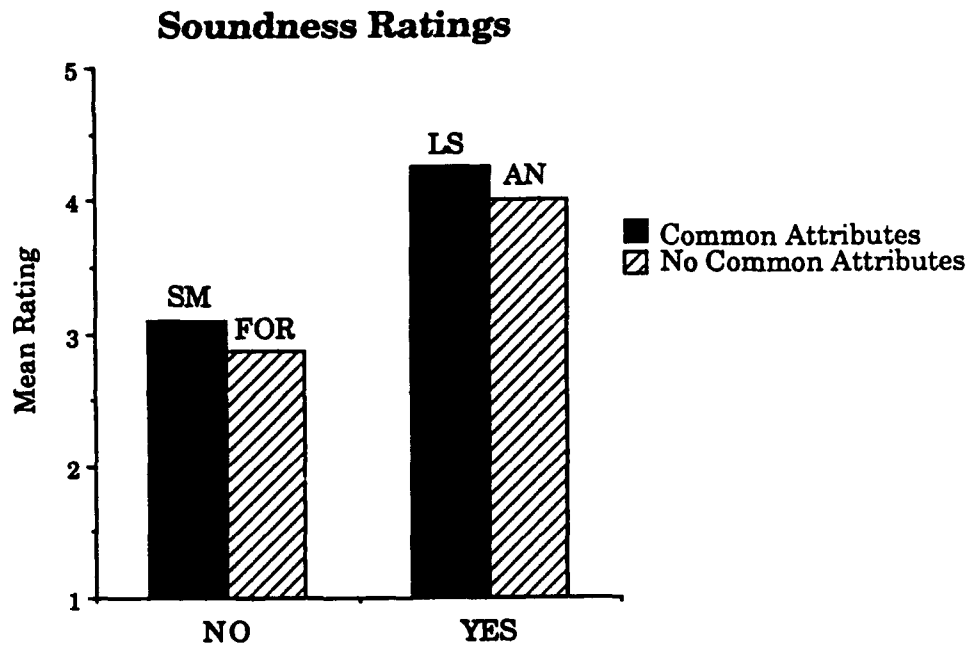
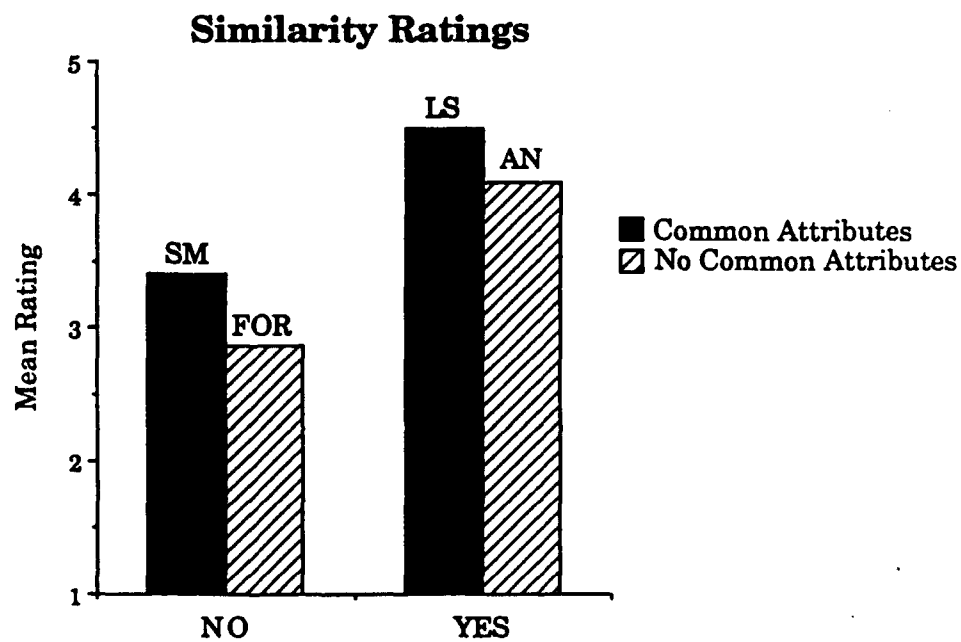


Figure 3a

Mean similarity rating for the four similarity types in Experiment 2



Common Higher-Order Relational Structure

Figure 3b

## Experiment 2b

In Experiment 2b, we tested the memory accessibility of the four kinds of similarity matches rated in Experiment 2a, using a cued recall task similar to that used in Experiment 1. After the recall task, these subjects rated the soundness and similarity of the story pairs. These ratings are subject to carryover, since the tasks were carried out in a fixed order. However, the results are of some interest as a within-subjects comparison of accessibility, soundness, and similarity.

### Method

#### Subjects

The subjects were 36 undergraduates who received class credit for participation. In addition, 18 paid subjects were used to rerun one cell in the soundness-rating task and one cell in the similarity-rating task due to experimenter error.

#### Materials and Design

The materials were identical to those used in Experiment 2a.

#### Procedure

Reminding. The procedure for the reminding task was the same as in Experiment 1. In the first session, subjects were told to read and remember 20 original stories and 12 filler stories. One week later, they returned for the reminding session and were given booklets of 20 cue stories, each of which matched one and only one original story. The instructions for the reminding task were identical to those of Experiment 1.

The data were scored as in Experiment 1, with two blind judges rating each recall on a 0-5 scale for accuracy to the original story. The judges agreed 79% of the time and were within one rating point of each other 97% of the time. The proportion of reminders above criterion and the score for key sentence recall were computed as in Experiment 1.

Soundness-rating task and similarity-rating task. As in Experiment 1, the soundness-rating task was administered after the reminding task. The instructions were the same as in Experiment 2a. Following the soundness-rating task, subjects were again given the same pairs of stories and asked to rate them on their overall similarity, again using the same instructions as in Experiment 2a.

## Results and Discussion

#### Reminding

Figure 4 shows the proportion of reminders (i.e., the proportion of recalls receiving scores above criterion) for the four kinds of similarity matches. As in Experiment 1, object commonalities contributed strongly to memory access, and higher-order relational commonalities had little effect. A 2 x 2 Relational similarity (within) x Object similarity (within) analysis of variance performed on the number of reminders revealed a main effect of Object similarity,  $F(1,35) = 108.73$ ,  $p < .0001$ . There was no effect of Relational similarity,  $F(1,35) = .89$ . Here,

as throughout the reminding results, there was no interaction between Object similarity and Relational similarity. Planned comparisons showed that the grouping of LS ( $M = .56$ ) and SM ( $M = .53$ ) was significantly different from the grouping of AN ( $M = .12$ ) and FOR ( $M = .09$ ). An item analysis also revealed a main effect of Object similarity,  $F(1,19) = 78.45$ ,  $p < .0001$ .

Proportion recalled for the four similarity types in Experiment 2

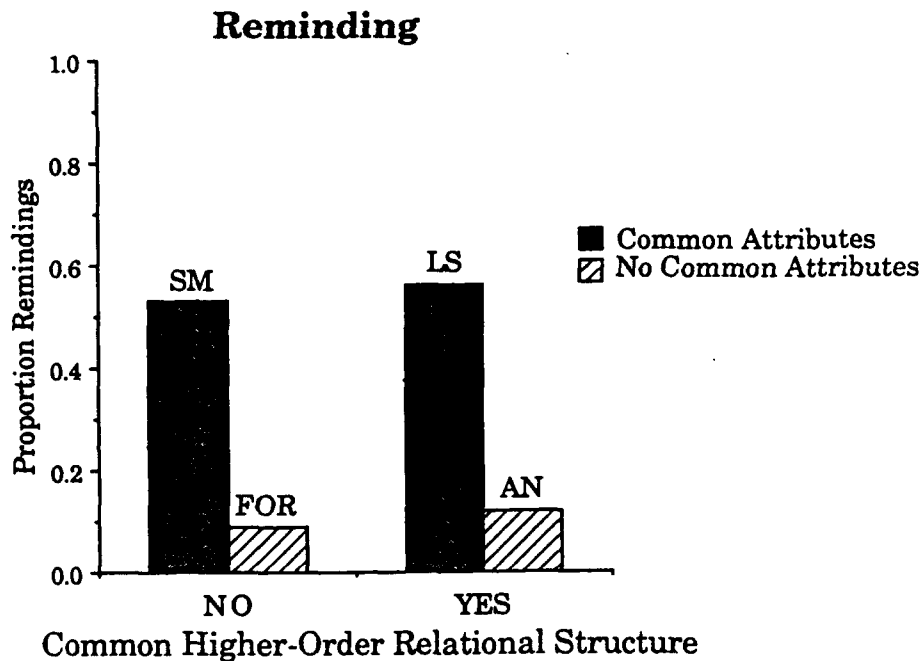


Figure 4

Similar analyses were performed on the overall quality-of-recall scores and the key word data. Table 4 shows all three measures of reminding – proportions of reminders, mean ratings of quality of recall, and mean key word scores across the three similarity match types. For comparison, the mean ratings of soundness and similarity from Experiment 2a are also included.

The analysis of the quality-of-reminding data revealed a main effect of Object similarity,  $F(1,35) = 124.18$ ,  $p < .0001$ , and also a main effect of Relational similarity,  $F(1,35) = 4.98$ ,  $p < .05$ . Planned comparisons showed a significant advantage for LS matches ( $M = 1.92$ ) and SM matches ( $M = 1.64$ ) over AN matches ( $M = .44$ ) and FOR matches ( $M = .27$ ). The item analysis also revealed a main effect of Object similarity,  $F(1,19) = 77.82$ ,  $p < .0001$ , and a main effect of Relational similarity,  $F(1,19) = 5.29$ ,  $p < .01$ .

The Key word analysis revealed main effects of Object similarity,  $F(1,35) = 11.86$ ,  $p < .01$ , but not Relational similarity,  $F(1,35) = .5$ . Although planned comparisons revealed no significant differences among the means, the ordering was similar to that obtained in the other two measures: LS matches ( $M = 1.11$ ), SM matches ( $M = .81$ ), FOR matches ( $M = .25$ ), and AN ( $M = .17$ ) matches. An item analysis also confirmed a main effect of Object similarity,  $F$

(1,19) = 7.45,  $p < .05$ , but not Relational similarity,  $F(1,19) = .646$ .

**Results of Experiment 2:  
Mean Ratings of soundness and similarity compared to  
three measures of reminding across the four similarity types.**

**Table 4**

Measure	Match Type			
	Literal Similarity (OA + R1 + R2) <sup>a</sup>	Analogy (R1 + R2)	Surface Matches (R0 + R1)	First-order Matches (R1)
Soundness <sup>b,c</sup>	4.26	4.01	3.10	2.87
Similarity	4.50	4.09	3.40	2.88
Proportion Recalled	.56	.12	.53	.09
Proportion of Key Words Recalled	1.11	.17	.81	.25
Quality of Reminding	1.92	.44	1.64	.27

Notes:

<sup>a</sup>Here OA refers to object commonalities, R1 to first-order relational commonalities, and R2 to higher-order relational commonalities.

<sup>b</sup>Soundness and similarity were rated by two independent groups of subjects.

<sup>c</sup>Soundness, similarity and quality of reminding are mean ratings on 1 (low) to 5 (high) scales.

The results of the reminding task replicate and extend the pattern of object-dominance in reminders found in Experiment 1. Matches that shared object-level commonalities – LS and SM matches – produced many more reminders than AN and FOR matches. In contrast,

higher-order relational similarity did not lead to a significant recall advantage in any of the three measures except quality of reminding, and this measure is not a pure estimate of memory access and indeed seems likely to overestimate the importance of relational similarity in memory. (This is because, as discussed above, any tendency subjects may have to use the cue story as a template for their written recall will tend to lead to improved quality-of-reminding scores for LS and AN (the two relationally similar match types) but not for MA and FOR.) Moreover, even under the quality of reminding measure, AN matches were not significantly better than FOR matches in access. The other two measures of recall, key word recall and proportion recalled, show effects of object similarity but not of higher-order relational similarity.

### Within-subject Ratings of Soundness and Similarity

The results of the soundness-rating task are similar to those of Experiment 2a in showing a strong effect of higher-order relational commonalities on subjective soundness. However, in contrast to Experiment 2a, object-level commonalities also contributed to soundness, possibly because of carryover from the recall task in which (as discussed above) object similarities were the major contributor to accessibility. A 2 x 2 Relational similarity (within) x Object similarity (within) analysis of variance revealed a main effect of both Relational similarity and Object similarity,  $F(1,35) = 159.76, p < .0001$  and  $F(1,35) = 10.91, p < .01$ , respectively. Bonferroni planned comparisons revealed a significant advantage of LS matches ( $M = 4.35$ ) over AN matches ( $M = 3.96$ ), AN matches over SM matches ( $M = 3.11$ ), and SM matches over FOR matches ( $M = 2.68$ ). An analysis over items also revealed a main effect of Relational similarity,  $F(1,19) = 55.74, p < .0001$ , and a main effect of Object similarity,  $F(1,19) = 15.11, p < .0001$ .

The similarity ratings were analyzed in the same way. Consistent with the results of Experiment 2a and with the predictions of structure-mapping, there were main effects of Relational similarity,  $F(1,35) = 130.58, p < .0001$ , and Object similarity,  $F(1,35) = 13.72, p < .0001$ . Planned comparisons revealed an advantage for LS matches ( $M = 4.56$ ) over AN matches ( $M = 4.00$ ) and for AN matches over SM matches ( $M = 3.10$ ). The difference between SM matches and FOR matches ( $M = 2.75$ ) was not significant. The item analysis revealed a main effect of Relational similarity,  $F(1,19) = 55.05, p < .0001$ , and a main effect of Object similarity,  $F(1,19) = 17.28, p < .001$ .

Despite the obvious potential for carryover from recall to soundness to similarity, the within-subject ratings are strikingly similar to those obtained from independent subjects in Experiment 2a. The fact that subjects gave low soundness ratings and even low similarity ratings to the very matches they themselves had just recalled is compelling evidence that different kinds of similarity figure in different cognitive tasks, as discussed below.

### Comparisons across tasks

As in Experiment 1, the results of the reminding task contrast sharply with those of the soundness-rating task. As before, the correlation between soundness<sup>10</sup> and proportion reminding was nonsignificant,  $r(78) = .008, NS$ . As predicted by structure-mapping, common systematic relational structure appears paramount in subjects' judgments of inferential soundness. Yet, it seems surprisingly unimportant in memorial access. Instead, similarity-based access appears strongly influenced by surface similarity, with common relational structure playing a smaller (though not negligible) role.

<sup>10</sup> All figures for soundness, unless otherwise stated, refer to the ratings done by independent raters.

However, before embracing this interpretation, we must ask whether overall similarity could have predicted these results. It is possible that the results of the reminding task simply reflect subjects' perceived overall similarity between the original and the cue. To the extent that this explanation holds, the reminding results should parallel the similarity ratings. As can be seen in Figures 3b and 4, the results do not bear out this possibility. The results of the reminding task show a very different pattern from those of the similarity-rating task. Indeed, the pattern for similarity resembles that for soundness more than it does that for recall, and the correlations across the 80 story matches bear out this impression. Similarity is highly correlated with soundness,  $r(78) = .74$ ,  $p < .01$ , but not with proportion recalled,  $r(78) = .17$ , NS. The finding that subjects are sensitive to common relational structure in judging similarity is consistent with several other recent findings. We return to this point in the discussion.

Subjects consider relational matches more *similar* than object matches yet *recall* more object matches than relational matches. These results suggest that different kinds of similarity govern access to long-term memory, on the one hand, and the processes of aligning, mapping, and evaluating similarity comparisons for which both members are already given on the other hand. This dissociation would appear to have serious consequences for models of human memory that rely on a unary notion of similarity (e.g., Gillund & Schiffrin, 1984; Hintzman, 1984; Medin & Schaffer, 1978 (but see Murphy & Medin, 1985)). In general, the pattern of results implies a more complex view of similarity than the one that currently prevails. We will return to this point in the Discussion.

### Experiment 3

In Experiments 1 and 2, we found that surface matches were highly accessible in memory. Specifically, we found that adding object commonalities to stories that shared first-order relations was much more effective in promoting recall than adding higher-order relations. This suggests that similarity-based access to memory may be particularly sensitive to common object descriptions. In Experiment 3, we tested this possibility by asking whether common object-descriptions by themselves would promote access. We created *object-only matches* – pairs of stories in which there were common object-descriptions but virtually no relational commonalities either in first-order events or in higher-order causal structure. Table 5 shows a sample pair: a study story (the same as for the prior experiments) along with its object-only cue story.

The same cued-recall method was used as in the prior studies to compare the accessibility of object-only matches (OM), which shared only object attributes with the base, with surface matches (SM), which shared both object attributes and first-order relations, and with analogy matches (AN), which shared first-order and higher-order relations. Aside from the question of access, another question that motivates the study is how the soundness and similarity for the two kinds of surface matches will compare with each other and with those for analogy. The prediction from structure-mapping, based on the degree of common relational structure, is that soundness will be highest for AN matches ( $R1 + R2$ ), then SM matches ( $OA + R1$ ), and finally OM matches ( $OA$ ).<sup>11</sup> For similarity, the structure-mapping prediction is that subjective similarity should be a function of both relational commonalities and object commonalities. However,

<sup>11</sup>For clarity, we will use a notation expressing predicate level: OA = lower-order commonalities (object descriptions); R1 = first-order relational commonalities (relations applying to objects: e.g., events), and R2 = higher-order relations (relations applying to first-order relations: e.g., causal connections between events).

to test these predictions requires further assumptions about the number and degree of similarities on different levels. In the present case, there is one clear prediction, namely, that OM matches should be rated lower in similarity than SM matches, since they contain fewer commonalities. A further, contingent expectation, based on the results of Experiment 2, is that with these materials relational commonalities may outweigh object level commonalities in determining subjective similarity, in which case AN matches will be rated more similar than SM matches.

**Sample materials from Experiment 3, showing an  
original story and its objects-only match.**

**Table 5**

Original story

Two small countries, Bolon and Salam, were adjacent to a large, warlike country called Mayonia. Bolon decided to make the best of the situation by taking over Salam. Salam started looking for aid from other strong countries but soon Bolon succeeded in taking it over. Then victorious Bolon proposed to make a treaty with its warlike neighbor Mayonia. Bolon proposed to give Mayonia control over Salam in exchange for a guarantee that Bolon would remain independent. Mayonia responded by overrunning both Bolon and Salam. Bolon was so busy maintaining control of Salam, it could do nothing to stop Mayonia. Thereupon, Mayonia installed puppet governments in both Bolon and Salam, and took over the newspapers and radio stations.

Objects-only Match

Two weak nations, Lincoln and Moreland bordered each other. Both countries relied heavily on the tourist trade to keep their economies afloat. They competed with each other over which one of them would get the most tourists. Meanwhile, another nearby nation, Chad, had a very strong economy with a thriving tourist trade. Tourist cruises flocked into its harbors and planes full of visitors were constantly landing in its airport. Because if this Moreland tried to join forces with Chad in its new advertising campaign to entice still more tourists. Unfortunately a hurricane hit the coast and bankrupted all three nations.

## Method

### Subjects

The subjects were 52 undergraduates from the University of Illinois, who were fulfilling a course requirement.

### Materials and Design

The materials were similar to those of the previous two studies, except that only AN matches and the two kinds of surface matches were used. Each subject received seven analogy matches (AN) and seven superficial and the other half received surface matches of the same kind used in the prior experiments (SM matches, which shared first-order relations as well as objects). In addition, each subject received six LS matches (the same six across all subjects). This was done to give subjects some literally-similar reminders to anchor their responses. Within each group, subjects were further divided into two counterbalancing groups that differed as to which stories were presented in each similarity type. These manipulations formed a 2 x 2 Surface-match type (between) x Similarity type (within) design.

### Procedure

The procedure was the same as that used in Experiments 1 and 2. In the first session, the subjects read the 20 original stories and 12 filler stories. One week later, they were given 20 matching stories in the reminding task followed by either the soundness-rating task or the similarity-rating task. Half the subjects received each of the rating tasks. Two judges rated the subjects' recalls. The inter-judge reliability was 75.8%.

## Results and Discussion

### Soundness-rating task

Table 6 shows the results for soundness, similarity, and reminding.<sup>12</sup> As predicted, the soundness ratings reflected degree of relational overlap: AN matches were rated as most sound, followed by SM matches and then OM matches. A 2 x 2 (Surface-match type (between) x Similarity type (within)) analysis of variance revealed significant effects of Similarity type (analogies versus superficial matches) and Surface-match type (whether the superficial matches were OM or SM) ( $F(1,24) = 48.57, p < .0001$  and  $F(1,24) = 13.53, p < .001$ ). The AN matches ( $M = 3.87$ ) were rated as significantly more sound than the SM matches ( $M = 3.11$ ) by a planned linear contrast, and the SM matches as significantly more sound than the OM matches ( $M = 1.92$ ) by a planned comparison. Although LS matches were omitted from the analysis (because they served as fillers and did not vary between groups), their soundness rating ( $M = 4.23$ ) was high, as expected. There was an interaction between Surface-match type and Similarity type,  $F(1,24) = 9.17, p < .006$ , reflecting the fact that AN is better recalled than OM but less well recalled than SM. An item analysis confirmed main effects of Similarity type,  $F(1,11) = 46.13, p < .0001$ , and Surface-match type,  $F(1,11) = 30.23, p < .0001$ , and an interaction between Surface-match type and Similarity type,  $F(1,11) = 12.67, p < .01$ .

---

<sup>12</sup>The means for the literal similarity matches are included for comparison but were not included in the analyses.

**Results of Experiment 3:**  
**Mean ratings of soundness and similarity compared to three**  
**measures of reminding across the three similarity types.**

Table 6

Measure	Match Type			
	Literal Similarity (OA + R1 + R2)	Analogy (R1 + R2)	Surface Matches (OA + R1)	Objects-only (OA) <sup>a</sup>
Soundness <sup>b</sup>	4.17	3.87	3.11	1.92
Similarity	4.47	3.55	4.01	2.90
Proportion Recalled	.62	.07	.52	.17
Quality of Reminding	1.79	.22	1.39	.47

Notes:

<sup>a</sup>Here OA refers to object commonalities, R1 to first-order relational commonalities, and R2 to higher-order relational commonalities.

<sup>b</sup>Soundness, similarity, and quality of reminding are mean ratings on 1 (low) to 5 (high) scales.

### Similarity-rating task

The results of the similarity-rating task are shown in Table 6. As predicted, SM matches ( $M = 4.01$ ) were rated as more similar than OM matches ( $M = 2.90$ ). However, in contrast to the finding of Experiment 2, SM matches were considered more similar than AN matches ( $M = 3.55$ ). An analysis of variance revealed a main effect of Surface-match type,  $F(1,24) = 4.24$ ,  $p < .05$ , and an interaction between Surface-match type and Similarity type,  $F(1,24) = 21.21$ ,  $p < .0001$ . There was no main effect of Similarity type. A planned linear contrast confirmed the advantage of SM matches over AN matches, and a planned comparison confirmed that AN matches were rated significantly more similar than the OM matches. An analysis over items showed the same effects: a main effect of Surface-match type,  $F(1,11) = 6.84$ ,  $p < .05$ , and an interaction between them,  $F(1,11) = 26.6$ ,  $p < .0001$ .

### Reminding

This experiment was motivated by two questions. The first was whether the dissociation between memory accessibility and subjective soundness and similarity would extend to the match types used here. The answer appears to be yes. Table 6 reveals that, as in the previous studies, the reminding pattern differs markedly from the pattern for soundness and even from the pattern for similarity. The second motivation was to determine whether object commonalities by themselves would promote memory access. As Table 6 shows, OM matches ( $M = .17$ ) were considerably less well retrieved than SM matches ( $M = .52$ ) (significant by a planned comparison). However, OM matches are better recalled than AN matches ( $M = .07$ ), although this difference was only marginally significant.<sup>13</sup> As expected, the anchoring LS matches were highly retrievable ( $M = .62$ ).

An analysis of variance on the proportions of reminders showed significant effects of Surface-match type,  $F(1,50) = 16.10$ ,  $p < .0001$ , and Similarity type,  $F(1,50) = 55.86$ ,  $p < .0001$ , and an interaction between Surface-match type and Similarity type,  $F(1,50) = 16.86$ ,  $p < .0001$ . The item analysis showed the same pattern: There were main effects of Surface-match type,  $F(1,11) = 36.71$ ,  $p < .0001$ , and Similarity type,  $F(1,11) = 15.68$ ,  $p < .0001$ , and an interaction between Surface-match type and Similarity type,  $F(1,11) = 5.99$ ,  $p < .05$ .

Similar analyses were performed on the quality-of-recall data and the key word data, revealing highly similar patterns of results. (See Table 6.) For rated quality of recalls, there were main effects of Surface-match type ( $F(1,50) = 23.85$ ,  $p < .0001$ ) and Similarity type ( $F(1,50) = 70.83$ ,  $p < .0001$ ) as well as an interaction between Surface-match type and Similarity type,  $F(1,50) = 19.82$ ,  $p < .0001$ . The mean quality of recalls was significantly higher for SM matches ( $M = 1.39$ ) than for OM matches ( $M = .47$ ) and for OM matches than for AN matches ( $M = .22$ ). The significant advantage for OM matches over AN matches is especially telling, since the quality-of-recall measure should favor AN matches over OM matches. For AN matches, a subject could benefit from using the cue story as a template to guide a reconstruction of the original story. This method would fail with OM matches, whose scores are therefore purer reflections of memory. As expected, LS matches received high ratings ( $M = 1.79$ ). The item analysis also confirmed main effects of Surface-match type,  $F(1,11) = 32.00$ ,  $p < .0001$ , and Similarity type,  $F(1,11) = 69.83$ ,  $p < .0001$ , and an interaction between Surface-match type and Similarity type,  $F(1,11) = 19.51$ ,  $p < .001$ .

As in the prior experiments, the pattern of results was weaker for the keyword analysis. Only the effects of Surface-match type reached significance,  $F(1,50) = 4.23$ ,  $p < .05$ . It was also significant in the analysis over items,  $F(1,11) = 5.67$ ,  $p < .05$ . There was a significant advantage in keyword recall for SM matches ( $M = .54$ ) over AN matches ( $M = .08$ ) and OM matches ( $M = .08$ ) but no difference between the OM matches and the AN matches. The keyword score for LS matches ( $M = .25$ ) was lower than expected.<sup>14</sup>

### Summary

Experiment 3 was carried out, first, in order to discover more about the determinants of memorial access: in particular, to find out which aspects of the SM matches led to the high accessibility that we observed in the previous two experiments. The results of Experiment 3 suggest that the high recall of the SM matches in those experiments was not due simply to the

<sup>13</sup>The difference between the AN and OM means was .083, just short of the critical difference of .084.

<sup>14</sup>Recall that only six LS matches were used; thus, specific item effects could affect the results.

presence of object commonalities. If this had been the case, then the OM and SM matches would have been equally well recalled in Experiment 3. In fact, SM matches were recalled significantly more often than OM matches (i.e., OA + R1 is superior to OA). However, OM matches were recalled better than the purely relational AN matches. It appears that object-level commonalities, even by themselves, can compete with higher-order relations in accessibility.

The second aim of Experiment 3 was to further examine the dissociation between different similarity types in recall vs. subjective soundness. Indeed, as in the prior studies, the soundness results show a different pattern from the reminding results. The AN matches were rated as significantly more sound than the SM matches, which were in turn rated more sound than the OM matches. As in the prior experiments, this ordering reflects the degree of relational overlap from AN matches, which share higher-order as well as first-order relations, through SM matches, which share first-order relations and object descriptions, and finally to OM matches, which share only object descriptions. This ordering contrasts sharply with the ordering of recallability: for example, analogy is rated as most sound but is least well recalled.

The results of the similarity ratings conform to the prediction that similarity should reflect both object similarity and relational similarity. Thus, both SM and AN matches are rated more similar than OM matches. There is no a priori reason to expect any particular similarity ordering between AN (R1 and R2) and SM (OA and R1), and indeed the results vary. In Experiment 2, AN matches were rated more similar than SM matches; the reverse was true in Experiment 3. However, despite this difference, the same point can be made here as in Experiment 2: The relative similarity of different match types, as measured by the similarity ratings, does not account for the order of recall. Thus, AN matches are rated as substantially more similar than OM matches but OM matches are recalled as well or better (by the quality of reminding measure). (The proportion recalled was .17 for OM and .07 for AN, a marginally significant difference.)

### Computational Simulation: The Structure-Mapping Engine

The results we have presented support the claim that relational commonalities play a special role in the computation of inferential soundness. Here we describe the results of running a computer simulation of structure-mapping on the materials from the previous experiment. We begin by reviewing the desiderata for a computational model of analogical soundness. Then we lay out how the simulation works and describe the results of running the simulation on the experimental materials. Finally, we speculate on extensions of the model to account for the memory recall phenomena and to model the computation of overall similarity.

A computational model of soundness must satisfy at least two criteria. The first, as demonstrated in the current research, is that the computation must be sensitive to both systematicity and structural consistency. That is, the comparison process must be sensitive not only to common relations between objects but also to common relations between relations (systematicity). Moreover, at both levels it is not enough merely to match lists of the relations present, but there must also be sensitivity to relational binding (structural consistency). For example, consider the following three possible analogs:

- (1) Event A causes Event B which causes Event C.
- (2) Event a causes Event b which causes Event c.
- (3) Event a causes Event b; Event c causes Event b.

Although all three situations have the same two higher-order relations between the first-order events, situations (1) and (2) are analogous but situations (1) and (3) are not. It is not sufficient for the same relations to occur in both analogs. To be sound, the match must map the relations between the relations in such a way as to preserve their relational bindings.

A second desideratum is that the system must be capable of focusing on relational commonalities without advance knowledge of the particular set of relations involved in a given similarity comparison. In the current research, subjects were given pairs of stories with the general instruction to 'interpret this comparison and judge its soundness.' There was no goal or contextual expectation that would have favored one set of relations over another, yet subjects displayed a highly ordered pattern of relational focusing, despite the fact that the relations of interest had to be derived from the comparison. Many models of analogy have postulated contextual focusing mechanisms that can act to drive the interpretation (e.g., mechanisms that focus on finding matches for key predicates in the target (Holyoak & Thagard, 1989a) or base (Forbus & Oblinger, 1990)). (See also Burstein, 1983; Carbonell, 1986; Holyoak, 1985.) But while some sort of contextual focusing mechanism may be appropriate for modeling certain goal-driven analogy uses<sup>15</sup> the present results (along with many others: e.g., Clement & Gentner, 1991; Goldstone, Medin, & Gentner, 1991) indicate that such a specific goal-focusing cannot be a necessary aspect of interpreting analogy or similarity comparisons. Indeed, a theory of analogy that required top-down expectations in order to interpret a comparison would be forced to bypass some of the most interesting similarity phenomena, namely the spontaneous noticing of similarity or analogy which then leads to (rather than resulting from) a heightened focus on features common to the two situations.

### The Structure-Mapping Engine

The Structure-Mapping Engine (SME) simulates the process of deriving a structurally consistent interpretation of an analogy or similarity comparison (Falkenhainer, Forbus, & Gentner, 1986, 1989/90). As part of its interpretation it includes a structural evaluation, which can be used to model the subjective ratings of inferential soundness (Skorstad, Falkenhainer, & Gentner, 1987; Forbus & Gentner, 1989).<sup>16</sup>

Given propositional representations of the base and target, SME starts by finding all possible relational identities. Specifically, for each relation in the base, it finds all identical relations in the target. These potential correspondences are called match hypotheses. Additional match

<sup>15</sup>In many cases the most parsimonious treatment of the influence of contextual goals on analogical interpretation is simply to assume that contextual goals influence (1) the initial representations of base and target input to the matcher and (2) the selection among competing interpretations, when there is more than one structurally consistent possibility.

<sup>16</sup>It also draws any further candidate inferences that would follow from the interpretations, although this will not concern us here.

hypotheses are created by checking the corresponding arguments of matched relations. If both arguments are entities or functions<sup>17</sup> SME creates a match hypothesis between them.

Initially, SME may have a large number of mutually inconsistent local matches. The next stage is to collect these local matches into global matches – systems of matches with consistent entity-pairings. This stage proceeds in several steps. First SME propagates correspondences between entities upwards through the match hypotheses to identify the largest connected collections of matched predicates with consistent object mappings. Since disconnected collections can sometimes be structurally consistent, SME then attempts to merge these collections to find the maximal sets of consistent match hypotheses. (In fact, merging such disconnected collections is the way the system derives candidate inferences.) These global matches, called Gmaps, are possible interpretations of the analogy.

Once Gmap construction is complete, a structural evaluation is computed for each global match. There are two components of this structural evaluation. First, a local evidence score is assigned to each of the constituent match hypotheses that make up the global match. Relational matches, which must be identical, receive the highest score (e.g., *pursue/pursue*); a slightly lower score is given to matches between identical functions (e.g., *size/size*); and a still lower score is given to matches between non-identical functions (e.g., *size/status*). This implements a preference for interpretations that maximize relational identity. The second component of structural evaluation increases the scores for matches that belong to common systems between base and target. The mechanism for this is a kind of 'trickle-down' process: Some portion of the evidence for a match hypothesis is passed down to the arguments of the relations (Forbus & Gentner, 1989). This implements a preference for systematicity.

### The Computational Experiment

Propositional representations were constructed for 9 of the 20 story sets used in Experiment 2. To achieve maximum contrast, only the base story, the analogy story, and the surface match from each set were represented. The stories were represented in three levels of richness of concrete detail (see Forbus & Gentner, 1989). However, all three levels yield the same basic results, so we show only one level – the richest in detail. The nine base-AN pairs and nine base-SM pairs were given to SME using its analogy rule set.<sup>18</sup>

The results are shown in Table 7, along with the human soundness results from Experiment 2. For each story set, we show SME's structural evaluation score (its measure of soundness) for the base-AN pair and for the base-SM pair. Two points are worth noting. First, SME's structural evaluation scores are higher throughout for the base-AN pair than for the base-SM pair. This of course follows from the fact that its evaluation algorithm favors common relational structure. Second, SME's pattern of preference fairly closely parallels the pattern for human subjects. In eight of the nine story sets, human subjects rated the analogy as significantly more sound than the surface match, and in all of these, SME assigned a higher evaluation to the analogy pair than to the surface match. For story set 19, subjects showed no significant

<sup>17</sup> Functions are used to represent dimensions such as *size* and *brightness* (see Falkenhainer, Forbus, & Gentner, 1989/90).

<sup>18</sup> As discussed below, SME can also be run as a literal similarity matcher, in which case it matches object attributes as well as relational structure. SME can also operate as a surface-similarity matcher, in which case it focuses only on object attributes.

preference, and SME's analogy advantage was particularly low for this story set. (However, SME's analogy advantage was also low for story set 8, where human subjects showed a decided preference for analogy.)

**Comparison of SME's structural evaluation scores  
with human soundness ratings**

**Table 7**

Story #	SME's Structural Evaluation Score			Human Subject's Soundness Ratings		
	AN	SM	AN>SM?	AN	SM	AN>SM?
5: Karla, hawk	14.2	7.1	+	4.2	3.0	*
7: Julius, mule	16.1	11.7	+	3.8	2.0	*
8: Percy, squirrel	8.7	8.0	+	4.0	2.8	*
9: steak, dog	17.9	8.0	+	4.8	2.8	*
10: Boris, business	20.6	15.1	+	3.6	2.2	*
13: Morris, prisoner	28.5	18.7	+	4.8	2.0	*
15: Fred, shepard	31.7	24.6	+	4.4	2.6	*
17: pioneers, divide	14.1	12.5	+	4.4	3.2	*
19: cobra, Pierre	29.2	28.0	+	4.0	4.2	NS

**Notes:**

a. These figures are for SME in analogy mode. The representations used here are rich in low-order concrete details.

b. + indicates that SME's structural evaluation score was greater for the analogy (AN) pair than for the surface-match (SM) pair.

c. \* indicates significant difference by a t-test.

Not surprisingly, there is a significant positive correlation between SME's structural evaluation scores and subjects' soundness ratings,  $r(16) = .46$ . As expected, there is no correlation between SME's structural evaluation scores and the proportion recalled by our subjects,  $r(16) = .01$ , nor are SME's scores significantly correlated with human similarity ratings,  $r(16) = .29$ .

SME's process of computing the match has many attractive features for modeling human processing. It begins blindly, and the emergence of clever or deep interpretations results from its ability to utilize interconnectivity between its initial local matches. It arrives at its interpretation (i.e., its highest-rated Gmap) by finding the most systematic mappable structure consistent with the 1-1 mapping rule. This method does not need to rely on a top-down, context-dependent processing rule, such as "Find the match that contains a current goal" (Holyoak, 1985) nor is it restricted to one particular kind of relational content, as in a rule like "Promote matches containing casual relations" (Winston, 1982). The systematicity principle operates to promote all and any systems of relations that include causal chains, goal structures, and other constraining relations.

Because it simply relies on matches among systems of elements, SME's technique of structural alignment can be extended straightforwardly to modeling ordinary similarity, by including in the match object attributes as well as relations and functions. One limitation is that SME's interpretations are extremely sensitive to its representations of base and target. It perceives similarity and analogy only when the two representations it is comparing contain at least some explicit identities. Thus, it would fail to see a perfectly acceptable match if the two situations happened to be represented in different vocabularies. We speculate that this limitation may be psychologically apt: that people may fail to notice potential matches, even when both comparands are present, if they are not represented in a comparable manner.

## Discussion

In this research we asked how different kind of similarities affect different aspects of the transfer process. We began by noting that similarity-based transfer can be decomposed into *accessing* a similar item<sup>19</sup> in memory, followed by processes of *mapping*, *evaluation*, *inference*, and (possibly) *extracting common structure*. Here we focus on access and evaluation, asking what determines which matches people *get* from memory, and what determines which matches they *keep*. We found a rather marked dissociation. The *accessibility* of matches from memory was strongly influenced by surface commonalities and less strongly influenced by structural commonalities. In contrast, the subjective *soundness* of comparisons was high for pairs that shared higher-order relational structure and low for pairs that did not. Finally, both kinds of commonalities contribute to subjects' sense of overall similarity, which emerges as a complex combination of both surface and structural commonalities.

We begin by summarizing the findings for soundness and for accessibility across the three studies. We then turn to similarity itself and the implications of these findings for theories of similarity. Finally, we discuss implications of this pattern for theories of similarity-based transfer.

<sup>19</sup> Although the current studies involved storage and access of individual cases, many of the same considerations can apply to generalizations or abstract rules.

### Subjective Soundness

The results are consistent with the prediction of structure-mapping theory that the subjective soundness of a similarity match is determined by the degree to which the analogs share relational structure. Three aspects of the results lends support to this prediction. First, all else being equal, adding common relations increases the perceived soundness of a match. In all three studies, adding higher-order relations to a first-order event match increased the subjective soundness of the comparison; that is, analogies (R1 + R2) were considered more sound than first-order event matches (R1). In Experiment 2, literal similarity matches (OA + R1 + R2) were considered more sound than surface matches (OA + R1). Finally, in Experiment 3, soundness was higher for surface matches (OA + R1) than for objects-only matches (OA); that is, the addition of first-order relational matches (OA + R1) to object-only matches (OA) increased soundness.

The second finding, and one that constitutes strong evidence for structural specificity in soundness judgments, is that the addition of object-attribute commonalities *fails* to increase soundness. The conditions for this test are met in Experiment 1 and twice in Experiment 2, and in all applicable cases, the prediction is borne out. In Experiment 1, surface matches (OA + R1) were considered no more sound than first-order event matches (R1). In Experiment 2, there are two instances of the relevant comparison, and in neither case do additional object-attribute commonalities result in higher soundness ratings. That is, subjective soundness was not higher for surface matches (OA + R1) than for FOR matches (R1), nor for literal similarity matches (OA + R1 + R2) than for analogies (R1 + R2).<sup>20</sup> A final piece of evidence is that the soundness of analogical matches (R1 + R2) is higher than that of surface matches (OA + R1). That is, beginning with first-order commonalities, adding higher-order commonalities is worth more, in terms of soundness, than adding object-attribute commonalities. Overall, the results indicate that the addition of common relations increases subjective soundness, but the addition of common object-attributes does not. However, the effect is still to prefer deeper structures.

These results suggest a scheme for computing similarity that gives more weight to higher-order relations than to lower-order predicates.<sup>21</sup> We might ask whether, instead of differentially weighting relations, these results could be accounted for more simply with a model in which soundness is a simple additive function of the number of feature matches, and the relational levels happen to contain many more features than the object-attribute level (so that analogical matches would share more features than surface matches). But this assumption is rendered implausible by the accessibility results. As discussed below, surface matches showed an advantage over analogical matches in memorial access. The claim that analogies share more features, but are less retrievable, than surface matches would be awkward at best.

These results are compatible with prior findings indicating that adults consider metaphors more apt when they can find relational interpretations of the metaphors (Gentner, 1988b; Gentner & Clement, 1988). In this research, subjects' ratings of the aptness of metaphors were positively correlated with the degree of relationality of their interpretations (as judged by

<sup>20</sup>These results are for independent raters. The soundness ratings obtained from the recall subjects did show some effect of object commonalities. However, as discussed above, these ratings were subject to carryover from the prior access task, which may have inflated the ratings for matches sharing surface commonalities (since those were what subjects best recalled).

<sup>21</sup>In fact, as discussed above, SME's structural evaluator achieves its preference for deep structure by using a trickle-down algorithm, in which a portion of the evidence for a match is passed down to its arguments.

independent judges) and either not correlated or negatively correlated with the degree of attributionality of their interpretations. Finally, in the current study, there was a strong positive correlation between our subjects' soundness ratings and SME's structural evaluation scores. Since these scores are specifically designed to reflect the degree of structural match (including the depth of the matching relational structure), this correlation is further evidence that the human perception of inferential soundness is specifically sensitive to the degree of common relational structure. Evaluating the soundness of a comparison is a highly selective process.

### Similarity-based Access

The results for access were almost the reverse of the results for soundness. Although subjects rated true analogies as being most sound, they tended not to retrieve true analogies during the reminding task. Instead, they were most likely to access surface matches. In all three experiments, adding common object attributes to a match increased the proportion retrieved: that is, recall of surface matches (OA + R1) was always greater than recall of FOR matches (R1) and recall of literal similarity matches (OA + R1 + R2) was greater than recall of analogical matches (R1 + R2). A stronger result in the same direction is that adding common object attributes increased accessibility more than adding common higher-order relations: that is, recall of surface matches (OA + R1) exceeded that of analogical matches (R1 + R2).<sup>22</sup>

Another important line of evidence comes from comparing the proportion recalled across various match types with SME's scores when run in different match modes. As discussed above, running SME in different modes provides us a way of obtaining the degree of match given various assumptions about which matches are used to form the overall interpretation. The proportion recalled correlates strongly with SME's surface-match scores but not with its structural match scores (its soundness scores: i.e., its scores when run as a 'pure analogy' matcher). These results suggest that superficial similarities, including object descriptions, play an important role in access.

### Dissociation between Access and Evaluation

This pattern of low accessibility for purely structural similarities and high accessibility for superficial similarities has been found by other researchers (Gilovich, 1981; Holyoak & Koh, 1987; Ross, 1984, 1987, 1989; Schumacher & Gentner, 1987, in preparation; Seifert, McKoon, Abelson, & Ratcliff, 1986). For example, Ross (1987) found that subjects' reminders to prior problems while trying to solve current problems are strongly affected by superficial similarities between the problems; in contrast, their ability to successfully use prior problems, once retrieved or presented by the experimenter, is governed by structural similarity rather than by surface similarity.<sup>23</sup> This pattern is consistent with the present findings, since subjective soundness is

<sup>22</sup>The proportions of surface matches retrieved across the three studies were .78, .52, and .53, respectively, substantially higher than the retrieval rate for analogies (.44, .12, and .07, respectively). Even though the subjects in Experiment 1 outperformed the subjects in Experiments 2 and 3, the same pattern of results was found.

<sup>23</sup>One interesting exception is that the nature of the object-correspondences between the two problems, which might be considered an aspect of surface similarity, affects use as well as access of prior analogs. We will return to this point below.

defined as the degree to which inferences from one analog can be applied in the other. Holyoak and Koh (1987) varied the degree of surface and structural similarities between a prior and a later problem, using four versions of the tumor/fortress problem. Based on the transfer results, they concluded that surface similarity had a large impact on retrieval of the source analog, while structural similarity affected both retrieval and application of the prior analog. Our results are consistent with these findings in suggesting that access is either broadly influenced by both surface and structural similarities (Holyoak, 1985) or disproportionately influenced by surface commonalities (Gentner, 1987, 1989).

However, the present results extend the findings of the problem-transfer studies by demonstrating a dissociation between the kind of similarity people can easily retrieve and the kind that they believe to be inferentially useful. The problem-transfer studies demonstrate that subjects often access superficially similar rather than structurally similar prior problems, even when they demonstrably benefit more from using the structurally similar prior problem. However, this does not tell us whether the subjects are aware of the inferential superiority of the structural match. It could be the case that their grasp of the domain principles in topics such as algebra or probability theory is inadequate to allow them to discern a structural match, but that they can still benefit if provided with one. Indeed, Reed (1987) provides evidence for exactly this kind of pattern. Using algebra word problems, Reed constructed four kinds of matches, roughly the same kinds as were used here. He found that subjects asked to judge the similarity of solution procedures between two problems (a task that should reflect chiefly structural similarity) were strongly influenced by surface similarity. In particular, they rated surface matches as either as high as (Experiment 2) or higher than (Experiment 1) analogies. However, when they were given pairs of problems and asked to use one to solve the other, they derived more benefit from analogous pairs than from surface-matched pairs. Given these results, the pattern of a surface-match advantage in *access* and a structural-match advantage in *use* clearly does not entitle us to conclude a deep processing dissociation. It might instead result simply from relatively weak structural knowledge.

The present studies address this question directly. First, the insufficient-knowledge explanation for the access results is implausible in the current research, since the domain principles were highly familiar social and physical causes. Secondly, and more importantly, the soundness ratings provide direct evidence that subjects' domain representations were sufficiently articulated to permit them to discern structural matches, *when both stories were present*. Across all three experiments, subjects uniformly rated analogies as substantially more sound than surface matches, yet recalled more surface matches than analogies. Similar results were reported by Schumacher and Gentner (1987; in preparation). Subjects were given lists of proverbs to remember and then cued with proverbs that were either structurally similar or surface-similar to the prior proverbs. Subjects experienced about twice as many surface reminders as structural reminders. Yet, when asked to rate the soundness and similarity of the individual pairs of proverbs, subjects uniformly rated the structural matches as more sound, and also more similar, than the surface matches. In this study, as in the current study, there was a marked dissociation between the matches that came to subjects' minds most easily and the matches they judged to be most useful in reasoning. This argues for a processing dissociation. Even when subjects' domain representations include structural information, different processes are responsive to different kinds of similarity, and in particular, similarity-based memory access

appears disproportionately sensitive to surface commonalities.<sup>24</sup>

### Effects of Domain Knowledge on Relational Access

Despite the gloomy picture painted so far, there is considerable evidence to suggest that relational access does occur (a) for experts in a domain and (b) when initial encoding of the study set is relatively intensive. Novick (1988) studied reminders for mathematics problems using novice and expert mathematics students. There were two prior problems for each of the test problems, one structurally similar and one superficially similar to the test problem. Experts were more likely than novices to retrieve the structurally similar problem. Faries and Reiser (1988) also used a similarity competition technique, and although their study was not designed as a novice-expert comparison, their results also bear on the issue of expertise. They taught subjects the programming language LISP in three two-day training sessions. On the third day, subjects received several target problems to solve, each of which was superficially similar to a prior problem, but was structurally similar to a different prior problem. With this degree of training, Faries and Reiser's subjects were able to access structurally similar problems despite the competing superficial similarities.<sup>25</sup>

The most common explanation for this novice-expert difference is that experts may be more likely to have explicit, accessible representations of the relational structures of the content domain. In contrast, novices either lack this knowledge or have it in an implicit, contextually embedded form. This domain representation view receives some support from the fact that analogical access is better when subjects are encouraged to induce an explicit schema during initial encoding.

Gick and Holyoak (1983) investigated this possibility by having subjects compare two prior analogs and write a summary of their commonalities. Forty-five percent of the subjects who compared two prior analogs solved the analogous problem without receiving a hint (compared to 30% for one prior analog, as discussed above). Interestingly, the goodness of the summarization was predictive of the likelihood of solving the target problem. Of the subjects who gave a good summary of the problem, 90% produced the problem solution without a hint. These findings support Holyoak's (1985) suggestion that intensive encoding of the relational structure can increase analogical access. Catrambone and Holyoak (1987) pursued the comparison methodology further. They found that subjects exhibited increased relational retrieval when they were required to compare two prior analogs, but not when they were simply given two prior analogs. In another manipulation of style of encoding, Schumacher & Gentner (1987) gave subjects a list of about 50 proverbs under one of two encoding tasks: writing out the meaning of each individual proverb or rating its cleverness. Subjects were then cued by proverbs that were either relationally similar or surface similar to the items on the study list. Analogous matches were recalled significantly better in the meaning condition than in the rating condition; surface matches were recalled equally well under both study conditions. Finally, Seifert, McKoon,

<sup>24</sup>However, there was evidence that higher-order relational similarities can also promote access: In Experiment 1 (but not in Experiment 2) analogies were retrieved more often than first-order relational matches.

<sup>25</sup>As in the problem-solving research discussed earlier, the Faries and Reiser method differs from the present method in an important respect. In the present task, subjects are told to write out *any* reminders that occur to them. In a problem-solving context, subjects may reject structurally inappropriate reminders without writing them down. This could lead to an underestimate of the degree to which superficial similarity affects initial access.

Abelson, and Ratcliff (1986) investigated priming effects in a sentence verification task between thematically similar (analogical) stories. In their first experiment, Seifert et al. failed to obtain any priming advantage between stories that shared plot themes over stories that did not. In another study, subjects first studied a list of themes and then judged the thematic similarity of pairs of stories. Under these conditions, a priming effect was obtained: When two stories were tested in sequence that had the same theme, a matching conclusion sentence was verified faster than when the two stories did not share the same theme. Seifert et al.'s pattern of results underscores both of the previous points. First, analogical similarities often go unnoticed. Second, conditions that promote structural encoding, such as asking subjects to compare two prior items, having subjects write out meanings, informing subjects in advance about the kinds of commonalities that may occur, or even using expert instead of novice subjects, tend to promote relational access.

### Analogical mapping

We have suggested a dissociation between the kind of similarity that leads to similarity-based *access* and the kind that enters into evaluating the subjective *soundness* (or inferential plausibility) of a similarity comparison. We could say that analogy begins stupid and ends smart. But what about the *mapping* process, the central process whereby two items are aligned and compared?<sup>26</sup> Is it governed solely by structural similarity, or do surface commonalities also play a role? A useful technique for exploring this issue is to put structural and object-level similarities in conflict by using a cross-mapping technique (Gentner & Toupin, 1986; Schumacher & Gentner, 1988b). Cross-mapping occurs when the correspondences dictated by object similarity are in conflict with those dictated by structural commonalities, as in the analogy

10::20 as 20::40.

The object similarity between 20 in the base and 20 in the target is in conflict with the structural mapping, which maps 10 onto 20 by virtue of their common relational roles. There is evidence that the ease with which the correct object correspondences can be made influences people's success or failure, at least in the initial mapping. Gentner and Toupin (1986) gave children aged six to nine a story-mapping task in which they had to retell a story with new characters. The children performed best when the match between the old story (the base) and the new story (the target) was one of literal similarity (i.e., high object similarities that were consistent with the relational correspondences). They performed next best with analogy (dissimilar objects in base and target), and they performed worst with cross-mapped analogy (object similarities that were inconsistent with the relational correspondences). Using a similar logic, Schumacher and Gentner (1988b) taught adults a simulated device model and then transferred the adults to another device. Adults required many fewer trials to learn the new device to criterion when it was literally similar to the initial device (i.e., similar-looking gauges played the same functional roles) than when the devices were cross-mapped (i.e., similar-looking

<sup>26</sup>According to the computational account given above, the order of processing in spontaneous similarity-based comparison is *access*, then *mapping*, then *evaluation*.

gauges played different functional roles). Ross (1989) investigated the effects of cross-mapping in access and use of a prior analog. He found that access to prior analogs was unaffected by cross-mapping. Access was promoted both by similarity in story lines and by similarity in the set of objects involved; however, it was not important whether the objects were in the proper correspondences. The pattern for use was quite different: When subjects (who had the correct formula) attempted to use a prior problem to help apply the formula, they were unaffected by superficial similarities in story lines, but they were affected by the nature of the object correspondences – they performed better with structurally consistent object correspondences than when the objects were cross-mapped. Taken in combination with the findings on evaluation, these results suggest that although the final evaluation of a comparison is governed chiefly by relational commonalities, object-level commonalities affect the ease or difficulty of achieving this structural interpretation.<sup>27</sup> Such a view is consistent with the computational model embodied in SME.

### Similarity

Concerning similarity itself, two aspects of the current results are noteworthy: first, the plurality of effects obtained – the fact that different kinds of similarity apparently participate differentially in different subprocesses – and second, the pervasive importance of structural commonalities in similarity judgments.

#### The plurality of similarity

Despite its centrality in theories of cognition, similarity is notoriously difficult to define. Ellis (1965) distinguished three methods of operationalizing similarity: (1) in terms of subjects' judgments, (2) in terms of variation along some physical scale, and (3) in terms of transfer itself, a method he considers 'most unsatisfactory.' The present results indicate that even within the rubric of transfer, further differentiation is necessary since it appears that different kinds of similarity participate differentially in different subprocesses. Subjects' judgments of soundness were governed by structural commonalities, their access to long-term memory was governed chiefly by surface commonalities, and their judgments of similarity were influenced by both structural and surface commonalities.

The dissociation between surface similarity and structural similarity is related to several recent discussions. Rips (1989) demonstrates a dissociation between similarity, typicality, and categorization. Murphy and Medin (1985) and Keil (1989) comment on the limited usefulness of simple similarity and point out that mere resemblance does not provide a sufficient basis for conceptual structure. While many natural categories, such as *deer* involve both surface similarity and structural similarity (that is, they enjoy relatively high literal similarity within their membership), there are other cases (e.g., *sharks and wolves as predators*) in which the grouping based on surface similarity is at odds with that based on structural similarity.<sup>28</sup> (Barsalou's ad hoc categories are another example.) In the current studies, judgments of soundness yielded

<sup>27</sup> This has led Ross (1989) to suggest that there are differential effects of different kinds of surface similarities.

<sup>28</sup> We would state this contrast as one of surface/overall similarity versus structural similarity rather than one of similarity versus conceptual structure. Quine (1969) made a similar distinction between 'brute similarity' and 'theoretical similarity.'

patterns quite distinct from the patterns that follow from surface similarity. Thus, different kinds of similarity matches can give rise to different groupings.

### Structural effects in similarity

There are several indications that subjective similarity judgments were sensitive to the presence of common relational structure. First, rather surprisingly, subjective similarity showed a pattern rather like that of subjective soundness. Similarity and soundness were highly correlated in Experiments 2 and 3.<sup>29</sup> Further, if we consider all cases where the comparison can be made, adding relational information increased subjective similarity in every case. Literal similarity matches (OA + R1 + R2) were considered more similar than surface matches (OA + R1) (Experiment 2); analogical matches (R1 + R2) were considered more similar than FOR-matches (R1) (Experiment 2); and surface matches (OA + R1) were considered more similar than object-only matches (OA) (Experiment 3).

Similarity is also increased by adding object commonalities: literal similarity matches are considered more similar than analogy matches and surface matches are considered more similar than FOR-matches (Experiment 2). Further, if we compare analogies (R1 + R2) with surface matches (OA + R1), we find that the former is rated as more similar in Experiment 2 and the latter in Experiment 3. Thus both structural and surface similarities contribute to subjective similarity.

There is evidence from simple perceptual tasks to support both claims: first, that common relational structure contributes to subjective similarity, and second, that there are multiple kinds of similarity. For example, Goldstone, Medin, and Gentner (1991) and Medin, Goldstone, and Gentner (1990) gave subjects similarity tasks involving triads of simple geometric figures. When asked to choose between a relational match and an attributional match to a standard, subjects generally preferred the relational match: e.g., XX would be considered more similar to OO than to XO.<sup>30</sup> Goldstone et al. also found support for the plurality of similarity claim. They present evidence that the subjective similarity of pairs of figures is decomposable into separable relational and attributional 'pools'; subjects appear to adopt either a relational bias or an attributional bias in rating similarity, depending on which kind of commonality already dominates (Goldstone's MAX rule) (Goldstone, Medin, & Gentner, 1991). Markman, Medin, and Gentner (in preparation) find that the effects of adding or subtracting common or distinctive low-order features to pairs of stimuli depend heavily on how the alterations affect the relational structure. The addition of perceptually distinct items that add a common relation – e.g., adding X and T to the pair <XO, TA> to make <XOX, TAT> – typically increases the similarity of the two scenes. However, the addition of perceptually similar or identical items that do not add a common relation – e.g., adding T to both numbers of the previous pair to give <XOT, TAT> – may not increase similarity and can even decrease it.

Further results indirectly supportive of the claim that relational structure contributes to similarity come from studies of perceptual discrimination (Pomerantz, Sager, & Stoeber, 1977;

<sup>29</sup>Similarity was not measured in Experiment 1.

<sup>30</sup>This preference is not invariable; indeed, the factors that affect the degree of relational preference are themselves worthy of study. They appear to include age (adults are more relational than children), the time permitted for the task (the relational preference is attenuated if subjects are forced to respond quickly), and blocking (subjects shift towards a more relational preferences across blocks of trials involving the same set of relations).

Lockhead & King, 1977). For example, Pomerantz, Sager, and Stoever (1977) found that subjects could more easily discriminate between the stimuli ) and ( when they were presented in the context of a third identical element, ), yielding the discrimination )) vs. (). Pomerantz et al. suggested that the effect of adding this contextual component was to promote emergent features, such as symmetry and intersection. To use our present terms, the added component introduces different relational structures in the two stimuli, thus making them more dissimilar and hence more discriminable. Supporting this emergent-relation hypothesis, Pomerantz et al. found a reverse configural effect for pairs such as () ( vs. () ). That is, subjects were less able to discriminate ( from ) in this configural context. In our terms, the added components introduce a new common relational structure (leaving the distinctive component isolated), thus making the figures more similar and hence less discriminable. These configural superiority (and inferiority) effects are readily interpreted in terms of the effects of adding distinctive (or common) relational structure on similarity.

Markman and Gentner (1990) gave subjects pairs of pictures that were deliberately constructed to contain cross-mappings: e.g., <XYZYX, voxov>. They found that when asked to say what the first x mapped onto in the second scene, subjects often chose the X. However, if subjects first performed a similarity judgment on the pair, they placed the objects in correspondence based on their position within the common relational structure: e.g., they would map X onto v in the above pair, rather than X onto x. This research underscores the importance of common relational structure in similarity, even perceptual similarity. It also demonstrates the plurality of similarity. For example, Markman and Gentner found that when the cross-mapped pairs were made more similar (or, in fact, identical, as in <<XYZYX, VOXOV>) people were much more likely to choose the object match.

Taken in combination, these results have important implications for theories of similarity. First, the dissociations found here suggest that we must decompose similarity in order to achieve an adequate account. The study of similarity in transfer has become the study of different classes of similarity in different subprocesses in transfer. Second, the finding of the importance of structural relations in subjective similarity suggests that a radically different account of the computation of similarity is required; one, which like the computational model (SME) discussed above, utilizes common structural connectivity among features rather than simple feature-set intersection or vector dot products.

At issue here is the very nature of similarity. Psychological models of similarity have been couched either in terms of mental distance (Shepard, 1984), in terms of common elements (e.g., Thorndike, 1903), or, in Tversky's (1977) elegant contrast model, in terms of common elements increasing similarity and distinctive elements decreasing it (Tversky, 1977; Tversky & Gati, 1982; Gati & Tversky, 1982). More recently, there has been a return to a kind of mental distance view in memory models, including many connectionist models, that take similarity to be a dot product over vector representations (e.g., Hintzman & Ludlam, 1980; Rumelhart, 1989). The present results indicate that structural considerations are integral to the subjective perception of similarity. These results indicate a general sensitivity to common structural connectivity in subjective similarity. Analogy may be the purest case of structural commonality, but common relational structure is also important in ordinary literal similarity, even in perceptual domains.

### Why surface access?

The present results imply a dissociation of similarity into structural and surface similarities, with different psychological privileges. In particular, they suggest that similarity-based access is relatively more governed by surface similarity than is similarity-based inference and subjective soundness. To put these findings in perspective, let us contrast them with the memory access patterns found in computational models of case-based reasoning. These models are concerned with how prior experiences are retrieved and used to bear on current problems. A striking feature of most of these models is their reliance on the use of structural similarities in retrieving prior examples. This retrieval can take many forms: e.g., searching for past plans that satisfy (or fail to satisfy) a similar goal (Carbonell, 1983; Hammond, 1986; Riesbeck, 1981) or accessing prior cases via a common abstraction or generalization (Greiner, 1988; Kass & Leake, 1987; Kass, Leake, & Owens, 1987; Kolodner, Simpson, & Sycara-Cyranski, 1985; Schank, 1982). For example, Hammond's (1986) program, CHEF, in the domain of Chinese cooking, searches for similar past plans (and their solutions) when current plans fail to achieve a desired goal. Kass, Leake, & Owens's (1987) program, SWALE, explains current events (such as the death of the race horse Swale) by retrieving and revising prior cases; it considers explanations such as a drug overdose, by analogy to Janis Joplin, or a bad heart, by analogy to Jim Fixx. The case-based reasoning emphasis on retrieving prior examples and generalizations that are inferentially useful seems perfectly reasonable, yet the human patterns found here do not appear to fit this scheme.

What should we make of this disparity? One conclusion might be that our human access processes are quite badly designed – vestigial leftovers, perhaps, from an earlier evolutionary stage. A second point of view, however, is the 'kind world' hypothesis (See Gentner, 1989; Medin & Ortony, 1989), namely, that in much of our experience surface information is strongly correlated with structural information. Things that look like dogs usually are dogs, though one might be fooled by a Tasmanian wolf. Perceptual information being plentiful and easy to process, it might be a rather good strategy for a large data base to take advantage of this correlation. This brings us to the second point: Human knowledge bases are vastly larger than any artificial base yet constructed. It may be that a structural indexing scheme would not scale up properly to human-memory size: e.g., we might find ourselves constantly swamped with relational reminders.<sup>31</sup> A third argument that perhaps surface access is not entirely suboptimal rests on a line of speculation concerning relational learning and the novice-expert shift, as discussed above. We begin by conjecturing that an important difference between novices and experts in a domain is that experts know the relational constructs for that domain (e.g., Chi, Feltovich, & Glaser, 1981), which enables them to show more relational access than novices (e.g., Novick, 1988). But if we further conjecture that the optimal relational constructs are domain-specific, and in fact differ considerably from one domain to another, then hasty abstraction could lead to costly and even dangerous errors. All this suggests that an adaptive learning course would be for novices to rely initially on highly conservative matches which rely heavily on surface commonalities (Forbus & Gentner, 1986; Gentner, 1989; Medin & Ross, 1989).

These considerations have led us to explore a two-stage model of similarity-based access, which we call MAC/FAC ("Many are called but few are chosen") (Gentner & Forbus, in prepa-

<sup>31</sup>This of course assumes that relational patterns are less distinctive than object constellations, a point which remains to be proved (but see Gentner, 1981, for some related arguments).

ration). Holyoak and Thagard (1989a) have explored an alternative organization of memory retrieval followed by mapping, using a connectionist implementation. We do not have space fully to describe our system here. Briefly, we set up a large base of representations of scenarios and then probe the system with a cue scenario. We use a 'stripped-down' version of SME to compare the probe to each example in memory; this first stage counts the total number of matches without enforcing structural consistency, and has the advantage of being quick. Then the set of winners is subjected to the full SME similarity-mapping process.<sup>32</sup> This two-tier process has some appealing features. Since the initial surface match does not require computing structural consistency, it is computationally cheap, which seems appropriate for this early part of the retrieval process. However, this economy has the disadvantage that the first stage sometimes fails to produce a legitimate structural match. In this case the system (like the subjects in our experiment) will be left with a best match from memory which it considers structurally unsound.

Finally, we close by noting a fascinating tension between the kind of primitive object-based similarity that seems to govern access and the kind of sophisticated, structural similarity that seems to govern mapping and evaluation. Quine (1969) speaks of 'brute similarity' and speculates that it is eventually supplanted by 'theoretical similarity.' (He goes on to suggest that both are supplanted by abstract conceptual structures (see Murphy & Medin, 1985), a development that does not concern us here.) Indeed, we believe there is considerable evidence for a shift in children's cognitive development from a reliance on overall similarity, to attribute-based similarity, and finally to relational similarity. The present results would suggest, however, that despite the adult ability to select sophisticated relational matches, the ancient brute is still with us in memory access.

---

<sup>32</sup>Recall that SME's similarity mapping process is highly sensitive to structural connectivity; it aims to capture literal similarity, not surface similarity. The stripped-down version of SME used to generate the initial memory-access set is similar to the surface match version discussed above.

## Appendix

### Instructions - Rating Soundness

Please read these instructions very carefully, and do not go on until you understand them.

This part of the experiment is about what makes a good match between two stories or situations. We all have intuitions about these things. Some kinds of resemblances seem important, while others seem weak or irrelevant. Here's an example:

Suppose you and a friend were having an argument about politics. And your friend brought up another situation, say, the way a car works, and said that the situations are fundamentally the same.

You might agree that his car example matches well with the political situation, and see the argument he is making. This is what we mean by a sound match. Or, you might reject his example; you might feel that although the situations look the same, the resemblance is actually just superficial. We call this a spurious match.

In this part of the experiment, we want you to use your intuitions about soundness - that is, about when two situations match well enough to make a strong argument.

On each page, we will show you a pair of stories. (You may recognize the stories because you have read them before. But don't worry about that.) You will probably notice that the two stories in each pair resemble each other, or 'match' in certain ways. We want you to rate how 'sound' the match between the two stories is. A sound match between two stories is one in which the essential aspects of the stories match. To put it another way, a sound match is one in which you can draw conclusions about the second story from the first. If the pair of stories match this way, give them a high rating. The opposite of a sound match is a spurious match. In a spurious match the resemblance between the stories is superficial. If the pair of stories is like this, give them a low rating.

The scale you will use looks like this:

THE MATCH BETWEEN THE TWO STORIES IS..... (circle one number)

extremely spurious	somewhat spurious	intermediate	fairly sound	extremely sound
1	2	3	4	5

In case you're still not quite sure what we mean by a sound match, here is another way to think about soundness: A sound match is strong enough that you can infer or predict much of the second story from the first. For example, suppose you read the first story and just the first part of the second story. Could you infer or predict, with fair accuracy, what happens in the rest of the second story? If you could predict it with reasonable accuracy, then this is a sound match. If you could not, then this is a weak, or spurious match.

Please keep these instructions in front of you so you can refer back to them whenever you need to.

## References

- Anderson, J. R., Farrell, R., & Sauers, R. (1984). Learning to program in LISP. *Cognitive Science*, 8, 87-129.
- Brown, A. L., Bransford, J. D., Ferrara, R. A., & Campione, J. C. (1983). Learning, remembering, and understanding. In P. H. Mussen (Ed.), *Handbook of child psychology: Vol. III. Cognitive development* (4th ed., pp. 77-166). New York: Wiley.
- Burstein, M. H. (1981). Concept formation through the interaction of multiple models. *Proceedings of the Third Annual Conference of the Cognitive Science Society* (pp. 271-273), Berkeley, CA. Hillsdale, NJ: Erlbaum.
- Burstein, M. H. (1983). Concept formation by incremental analogical reasoning and debugging. *Proceedings of the International Machine Learning Workshop* (pp. 19-25), Monticello, IL. Urbana, IL: University of Illinois.
- Carbonell, J. G. (1981). Invariance hierarchies in metaphor interpretation. *Proceedings of the Third Annual Conference of the Cognitive Science Society* (pp. 292-295), Berkeley, CA. Hillsdale, NJ: Erlbaum.
- Carbonell, J. G. (1983). Learning by analogy: Formulating and generalizing plans from past experience. In R. S. Michalski, J. G. Carbonell, & T. M. Mitchell (Eds.), *Machine learning: An artificial intelligence approach* (Vol. 1, pp. 137-161). Palo Alto, CA: Tioga Publishing.
- Carbonell, J. G. (1986). Derivational analogy: A theory of reconstructive problem solving and expertise acquisition. In R. S. Michalski, J. G. Carbonell, & T. M. Mitchell (Eds.), *Machine learning: An artificial intelligence approach* (Vol. 2, pp. 371-392). Los Altos, CA: Morgan Kaufmann.
- Catrambone, R., & Holyoak, K. J. (1987, May). *Do novices have schemas?* Paper presented at the meeting of the Midwestern Psychological Association, Chicago, IL.
- Chi, M. T. H., Feltovich, P. J., & Glaser, R. (1981). Categorization and representation of physics problems by experts and novices. *Cognitive Science*, 5, 121-152.
- Clement, C., & Gentner, D. (1988). Systematicity as a selection constraint in analogical mapping. *Proceedings of the Tenth Annual Conference of the Cognitive Science Society* (pp. 412-418), Montreal. Hillsdale, NJ: Erlbaum.
- Clement, C. A., & Gentner, D. (1991). Systematicity as a selection constraint in analogical mapping. *Cognitive Science*, 15, 89-132.
- Clement, J. (1986). Methods for evaluating the validity of hypothesized analogies. *Proceedings of the Eighth Annual Conference of the Cognitive Science Society* (pp. 223-234), Amherst, MA. Hillsdale, NJ: Erlbaum.
- Duncker, K. (1945). On problem-solving. *Psychological Monographs*, 58(5, Whole No. 270).
- Ellis, H. C. (1965). *The transfer of learning*. New York: Macmillan.
- Falkenhainer, B., Forbus, K. D., & Gentner, D. (1986). The structure-mapping engine. *Proceedings of the Fifth National Conference on Artificial Intelligence* (pp. 272-277), Philadelphia, PA. Los Altos, CA: Morgan Kaufmann.
- Falkenhainer, B., Forbus, K. D., & Gentner, D. (1989/90). The structure-mapping engine: Algorithm and examples. *Artificial Intelligence*, 41, 1-63.

- Faries, J. M., & Reiser, B. J. (1988). Access and use of previous solutions in a problem solving situation. *Proceedings of the Tenth Annual Conference of the Cognitive Science Society* (pp. 433-439), Montreal. Hillsdale, NJ: Erlbaum.
- Forbus, K. D., & Gentner, D. (1986). Learning physical domains: Toward a theoretical framework. In R. S. Michalski, J. G. Carbonell, & T. M. Mitchell (Eds.), *Machine learning: An artificial intelligence approach* (Vol. 2, pp. 311-348). Los Altos, CA: Morgan Kaufmann.
- Forbus, K. D., & Gentner, D. (1989). Structural evaluation of analogies: What counts? *Proceedings of the Eleventh Annual Conference of the Cognitive Science Society* (pp. 341-348), Ann Arbor, MI. Hillsdale, NJ: Erlbaum.
- Forbus, K. D., & Oblinger, D. (1990). Making SME greedy and pragmatic. *Proceedings of the Twelfth Annual Conference of the Cognitive Science Society* (pp. 61-68), Cambridge, MA. Hillsdale, NJ: Erlbaum.
- Gati, I., & Tversky, A. (1982). Representations of qualitative and quantitative dimensions. *Journal of Experimental Psychology: Human Perception and Performance*, 8, 325-340.
- Gentner, D. (1980). *The structure of analogical models in science* (Tech. Rep. No. 4451). Cambridge, MA: Bolt Beranek and Newman, Inc.
- Gentner, D. (1981). Some interesting differences between nouns and verbs. *Cognition and Brain Theory*, 4, 161-178.
- Gentner, D. (1982). Are scientific analogies metaphors? In D. S. Miall (Ed.), *Metaphor: Problems and perspectives* (pp. 106-132). Brighton, Sussex: Harvester Press.
- Gentner, D. (1983). Structure-mapping: A theoretical framework for analogy. *Cognitive Science*, 7, 155-170.
- Gentner, D. (1987). *Mechanisms of analogical learning* (Tech. Rep. No. UIUCDCS-R-87-1381). Urbana, IL: University of Illinois, Department of Computer Science.
- Gentner, D. (1988a). Analogical inference and analogical access. In A. Frieditis (Ed.), *Analogica* (pp. 63-88). Los Altos, CA: Morgan Kaufmann.
- Gentner, D. (1988b). Metaphor as structure mapping: The relational shift. *Child Development*, 59, 47-59.
- Gentner, D. (1989). The mechanisms of analogical learning. In S. Vosniadou & A. Ortony (Eds.), *Similarity and analogical reasoning* (pp. 199-241). New York: Cambridge University Press.
- Gentner, D., & Clement, C. (1988). Evidence for relational selectivity in the interpretation of analogy and metaphor. In G. H. Bower (Ed.), *The psychology of learning and motivation: Advances in research and theory* (Vol. 22, pp. 307-358). New York: Academic Press.
- Gentner, D., Falkenhainer, B., & Skorstad, J. (1988). Viewing metaphor as analogy. In D. H. Helman (Ed.), *Analogical reasoning: Perspectives of artificial intelligence, cognitive science, and philosophy* (pp. 171-177). Dordrecht, The Netherlands: Kluwer.
- Gentner, D., & Forbus, K. D. (in preparation). A computational simulation of a two-stage model of similarity in transfer.
- Gentner, D., & Landers, R. (1985). Analogical reminding: A good match is hard to find. *Proceedings of the International Conference on Cybernetics and Society* (pp. 607-613), Tucson, AZ. New York: IEEE.

- Gentner, D., & Schumacher, R. M. (1986). Use of structure mapping theory for complex systems. *Proceedings of the 1986 IEEE International Conference on Systems, Man, and Cybernetics* (pp. 252-258), Atlanta, GA. New York: IEEE.
- Gentner, D., & Toupin, C. (1986). Systematicity and surface similarity in the development of analogy. *Cognitive Science*, 10, 277-300.
- Gick, M. L., & Holyoak, K. J. (1980). Analogical problem solving. *Cognitive Psychology*, 12, 306-355.
- Gick, M. L., & Holyoak, K. J. (1983). Schema induction and analogical transfer. *Cognitive Psychology*, 15, 1-38.
- Gillund, G., & Shiffrin, R. M. (1984). A retrieval model for both recognition and recall. *Psychological Review*, 91, 1-67.
- Gilovich, T. (1981). Seeing the past in the present: The effect of associations to familiar events on judgments and decisions. *Journal of Personality and Social Psychology*, 40, 797-808.
- Goldstone, R. L., Gentner, D., & Medin, D. L. (1989). Relations relating relations. *Proceedings of the Eleventh Annual Conference of the Cognitive Science Society* (pp. 131-138), Ann Arbor, MI. Hillsdale, NJ: Erlbaum.
- Goldstone, R. L., Medin, D. L., & Gentner, D. (1991). Relational similarity and the non-independence of features in similarity judgments. *Cognitive Psychology*, 23, 222-262.
- Greiner, R. (1988). Abstraction-based analogical inference. In D. H. Helman (Ed.), *Analogical reasoning: Perspectives of artificial intelligence, cognitive science, and philosophy* (pp. 147-170). Dordrecht, The Netherlands: Kluwer.
- Hall, R. P. (1989). Computational approaches to analogical reasoning: A comparative analysis. *Artificial Intelligence*, 39, 39-120.
- Hammond, K. J. (1986). CHEF: A model of case-based planning. *Proceedings of the Fifth National Conference on Artificial Intelligence* (pp. 267-271). Philadelphia, PA. Los Altos, CA: Morgan Kaufmann.
- Hintzman, D. L. (1984). MINERVA 2: A simulation model of human memory. *Behavior Research Methods, Instruments, & Computers*, 16, 96-101.
- Hintzman, D. L., & Ludlam, G. (1980). Differential forgetting of prototypes and old instances: Simulation by an exemplar-based classification model. *Memory & Cognition*, 8, 378-382.
- Holyoak, K. J. (1985). The pragmatics of analogical transfer. In G. H. Bower (Ed.), *The psychology of learning and motivation: Advances in research and theory* (Vol. 19, pp. 59-87). New York: Academic Press.
- Holyoak, K. J., & Koh, K. (1987). Surface and structural similarity in analogical transfer. *Memory & Cognition*, 15, 332-340.
- Holyoak, K. J., & Thagard, P. (1989a). Analogical mapping by constraint satisfaction. *Cognitive Science*, 13, 295-355.
- Holyoak, K. J., & Thagard, P. R. (1989b). A computational model of analogical problem solving. In S. Vosniadou & A. Ortony (Eds.), *Similarity and analogical reasoning* (pp. 242-266). New York: Cambridge University Press.

- Kass, A., & Leake, D. (1987). *A case-based approach to building explanations for explanation-based learning*. Unpublished manuscript, Yale University, Department of Computer Science, New Haven, CN.
- Kass, A., Leake, D., & Owens, C. (1987). SWALE, a program that explains. In R. Schank (Ed.), *Explanation patterns: Understanding mechanically and creatively*. Hillsdale, NJ: Erlbaum.
- Keane, M. (1985). On drawing analogies when solving problems: A theory and test of solution generation in an analogical problem-solving task. *British Journal of Psychology*, 76, 449-458.
- Kedar-Cabelli, S. (1988). Toward a computational model of purpose-directed analogy. In A. Prieditis (Ed.), *Analogica* (pp. 89-107). Los Altos, CA: Morgan Kaufmann.
- Keil, F. C. (1989). *Concepts, kinds, and cognitive development*. Cambridge, MA: MIT Press.
- Kolodner, J. L., Simpson, R. L., Jr., & Sycara-Cyranski, K. (1985). A process model of case-based reasoning in problem-solving. *Proceedings of the Ninth International Joint Conference on Artificial Intelligence* (pp. 284-290), Los Angeles, CA. Los Altos, CA: Morgan Kaufmann.
- Lockhead, G. R., & King, M. C. (1977). Classifying integral stimuli. *Journal of Experimental Psychology: Human Perception and Performance*, 3, 436-443.
- Markman, A. B., & Gentner, D. (1990). Analogical mapping during similarity judgments. *Proceeding of the Twelfth Annual Conference of the Cognitive Science Society* (pp. 38-44), Cambridge, MA. Hillsdale, NJ: Erlbaum.
- Markman, A. B., Medin, D. L., & Gentner, D. (in preparation). An analysis of the perceptual-verbal distinction in similarity judgments.
- Marr, D. (1982). *Vision: A computational investigation into the human representation and processing of visual information*. San Francisco: Freeman.
- Medin, D. L., Goldstone, R. L., & Gentner, D. (1990). Similarity involving attributes and relations: Judgments of similarity and difference are not inverses. *Psychological Science*, 1, 64-69.
- Medin, D., & Ortony, A. (1989). Psychological essentialism. In S. Vosniadou & A. Ortony (Eds.), *Similarity and analogical reasoning* (pp. 179-195). New York: Cambridge University Press.
- Medin, D. L., & Ross, B. H. (1989). The specific character of abstract thought: Categorization, problem-solving, and induction. In R. J. Sternberg (Ed.), *Advances in the psychology of human intelligence* (Vol. 5, pp. 189-223). Hillsdale, NJ: Erlbaum.
- Medin, D. L., & Schaffer, M. M. (1978). Context theory of classification learning. *Psychological Review*, 85, 207-238.
- Murphy, G. L., & Medin, D. L. (1985). The role of theories in conceptual coherence. *Psychological Review*, 92, 289-316.
- Novick, L. R. (1988). Analogical transfer, problem similarity, and expertise. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 14, 510-520.
- Ortony, A. (1979). Beyond literal similarity. *Psychological Review*, 86, 161-180.

- Osgood, C. E. (1949). The similarity paradox in human learning: A resolution. *Psychological Review*, 56, 132-143.
- Palmer, S. E., & Kimchi, R. (1985). The information processing approach to cognition. In T. Knapp & L. C. Robertson (Eds.), *Approaches to cognition: Contrasts and controversies* (pp. 37-77). Hillsdale, NJ: Erlbaum.
- Pirolli, P. (1985). *Problem solving by analogy and skill acquisition in the domain of programming*. Unpublished manuscript.
- Pomerantz, J. R., Sager, L. C., & Stoeber, R. J. (1977). Perception of wholes and of their component parts: Some configural superiority effects. *Journal of Experimental Psychology: Human Perception and Performance*, 3, 422-435.
- Quine, W. V. (1969). *Ontological relativity and other essays*. New York: Columbia University Press.
- Read, S. J. (1983). Once is enough: Causal reasoning from a single instance. *Journal of Personality and Social Psychology*, 45, 323-334.
- Read, S. J. (1984). Analogical reasoning in social judgment: The importance of causal theories. *Journal of Personality and Social Psychology*, 46, 14-25.
- Reed, S. K. (1987). A structure-mapping model for word problems. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 13, 124-139.
- Reed, S. K., Ernst, G. W., & Banerji, R. (1974). The role of analogy in transfer between similar problem states. *Cognitive Psychology*, 6, 436-450.
- Riesbeck, C. K. (1981). Failure-driven reminding for incremental learning. *Proceedings of the Seventh International Joint Conference on Artificial Intelligence*, Vancouver, B. C. Los Altos, CA: Morgan Kaufmann.
- Rips, L. J. (1989). Similarity, typicality, and categorization. In S. Vosniadou & A. Ortony (Eds.), *Similarity and analogical reasoning* (pp. 21-59). New York: Cambridge University Press.
- Ross, B. H. (1984). Reminders and their effects in learning a cognitive skill. *Cognitive Psychology*, 16, 371-416.
- Ross, B. H. (1987). This is like that: The use of earlier problems and the separation of similarity effects. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 13, 629-639.
- Ross, B. H. (1989). Reminders in learning and instruction. In S. Vosniadou & A. Ortony (Eds.), *Similarity and analogical reasoning* (pp. 438-469). New York: Cambridge University Press.
- Rumelhart, D. E. (1989). Toward a microstructural account of human reasoning. In S. Vosniadou & A. Ortony (Eds.), *Similarity and analogical reasoning* (pp. 298-312). New York: Cambridge University Press.
- Schank, R. C. (1982). *Dynamic memory*. New York: Cambridge University Press.
- Schumacher, R. M., & Gentner, D. (1987, May). *Similarity-based reminders: The effects of similarity and interitem distance*. Paper presented at meeting of the Midwestern Psychological Association, Chicago, IL.

- Schumacher, R. M., & Gentner, D. (1988a). Remembering causal systems: Effects of systematicity and surface similarity in delayed transfer. *Proceedings of the Human Factors Society 32nd Annual Meeting* (pp. 1271-1275), Anaheim, CA. Santa Monica, CA: Human Factors Society.
- Schumacher, R. M., & Gentner, D. (1988b). Transfer of training as analogical mapping. *IEEE Transactions on Systems, Man, and Cybernetics*, 18, 592-600.
- Schumacher, R. M., & Gentner, D. (in preparation). Analogical access: Effects of repetition, competition, and instruction.
- Seifert, C. M., McKoon, G., Abelson, R. P., & Ratcliff, R. (1986). Memory connections between thematically similar episodes. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 12, 220-231.
- Shepard, R. N. (1984). Ecological constraints on internal representation: Resonant kinematics of perceiving, imagining, thinking, and dreaming. *Psychological Review*, 91, 417-447.
- Simon, H. A., & Hayes, J. R. (1976). The understanding process: Problem isomorphs. *Cognitive Psychology*, 8, 165-190.
- Skorstad, J., Falkenhainer, B., & Gentner, D. (1987). Analogical processing: A simulation and empirical corroboration. *Proceedings of the Sixth National Conference on Artificial Intelligence* (pp. 322-326), Seattle, WA. Los Altos, CA: Morgan Kaufmann.
- Thorndike, E. L. (1903). *Educational psychology*. New York: Lemcke and Buechner.
- Tversky, A. (1977). Features of similarity. *Psychological Review*, 84, 327-352.
- Tversky, A., & Gati, I. (1982). Similarity, separability, and the triangle inequality. *Psychological Review*, 89, 123-154.
- Winston, P. H. (1980). Learning and reasoning by analogy. *Communications of the ACM*, 23, 689-703.
- Winston, P. H. (1982). Learning new principles from precedents and exercises. *Artificial Intelligence*, 19, 321-350.
- Wolstencroft, J. (1989). Restructuring, reminding, and repair: What's missing from models of analogy? *AI Communications*, 2, 58-71.

# Distribution List

Technical Document Center  
AFHRL/LRS-TDC  
Wright-Patterson AFB  
OH 45433-6503

Mr. Tejwani S. Anand  
Philips Laboratories  
345 Scarborough Road  
Briarcliff Manor  
New York, NY 10520

Dr. Nancy S. Anderson  
Department of Psychology  
University of Maryland  
College Park, MD 20742

Dr. Thomas H. Anderson  
Center for the Study of Reading  
174 Children's Research Center  
51 Gerry Drive  
Champaign, IL 61820

Dr. Michael E. Atwood  
NYNEX  
AI Laboratory  
500 Westchester Avenue  
White Plains, NY 10604

Dr. Alan Baddeley  
Medical Research Council  
Applied Psychology Unit  
15 Chaucer Road  
Cambridge CB2 2EF  
ENGLAND

Dr. James D. Baker  
Director of Automation and Research  
Allen Corporation of America  
209 Madison Street  
Alexandria, VA 22314

Dr. Meryl S. Baker  
Navy Personnel R&D Center  
San Diego, CA 92152-6800

Dr. Donald E. Bamber  
Code 446  
Naval Ocean Systems Center  
San Diego, CA 92152-5000

Dr. Isaac Bejar  
Law School Admissions  
Services  
P.O. Box 40  
Newtown, PA 18940-0040

Dr. John Black  
Teachers College, Box 8  
Columbia University  
525 West 120th Street  
New York, NY 10027

Dr. Deborah A. Boehm-Davis  
Department of Psychology  
George Mason University  
4400 University Drive  
Fairfax, VA 22030

Dr. Sue Bogner  
Army Research Institute  
ATTN: PERI-SF  
5001 Eisenhower Avenue  
Alexandria, VA 22333-5600

Dr. C. Alan Boneau  
Department of Psychology  
George Mason University  
4400 University Drive  
Fairfax, VA 22030

Naval Supply Systems Command  
NAVSUP 5512  
ATTN: Sandra Borden  
Washington, D.C. 20376-5000

Dr. Gordon H. Bower  
Department of Psychology  
Stanford University  
Stanford, CA 94306

Dr. Joanne Capper, Director  
Center for Research into Practice  
3545 Albemarle Street, NW  
Washington, DC 20008

CDR Robert Carter  
Office of the Chief  
of Naval Operations  
OP-933D4  
Washington, DC 20350-2000

Dr. Richard Catrambone  
Department of Psychology  
Georgia Institute of Technology  
Atlanta, GA 30332-0170

Dr. Ruth W. Chebey  
CDEC, Hamburg Hall  
Carnegie Mellon University  
Pittsburgh, PA 15213

Dr. W. Chambers  
Technology Manager, Code 2B  
Naval Training Systems Center  
12350 Research Parkway  
Orlando, FL 32826-3224

Dr. Davide Charney  
English Department  
Penn State University  
University Park, PA 16802

Dr. Michelene Chi  
Learning R & D Center  
University of Pittsburgh  
3939 O'Hara Street  
Pittsburgh, PA 15260

Mr. Michael Cowen  
Code 142  
Navy Personnel R&D Center  
San Diego, CA 92152

Dr. Meredith P. Crawford  
3563 Hamlet Place  
Chevy Chase, MD 20815

Dr. Hans F. Crombag  
Faculty of Law  
University of Limburg  
P.O. Box 616  
Maastricht  
The NETHERLANDS 6200 MD

Brian Dalkman  
Training Technology Branch  
3400 TCHTW/TTGXC  
Lowry AFB, CO 80230-5000

Dr. Antonio R. Damasio  
Department of Neurology  
College of Medicine  
University of Iowa  
Iowa City, IA 52242

Margaret Day, Librarian  
Applied Science Associates  
P.O. Box 1072  
Butler, PA 16003

Dr. Sharon Derry  
Florida State University  
Department of Psychology  
Tallahassee, FL 32306

Dr. David Diamond  
Department of Pharmacology  
Health Sciences Center  
University of Colorado  
P.O. Box C 236  
Denver, CO 80262

Defense Technical  
Information Center  
Cameron Station, Bldg 5  
Alexandria, VA 22314  
(2 Copies)

Dr. Ralph Dusek  
V-P Human Factors  
JIL Systems  
1225 Jefferson Davis Hwy.  
Suite 1209  
Arlington, VA 22201

Dr. Susan Epstein  
144 S. Mountain Avenue  
Montclair, NJ 07042

ERIC Facility-Acquisitions  
2440 Research Blvd, Suite 550  
Rockville, MD 20850-3238

Dr. Debra Evans  
Applied Science Associates, Inc.  
P. O. Box 1072  
Butler, PA 16003

Dr. Lorraine D. Eyde  
Office of Personnel Management  
Office of Examination Development  
1900 E St., NW  
Washington, DC 20415

LCDR Micheline Y. Eyraud  
Code 602  
Naval Air Development Center  
Warminster, PA 18974-5000

Dr. Jean-Claude Falmagne  
Irvine Research Unit in  
Mathematical & Behavioral Sciences  
University of California  
Irvine, CA 92717

Dr. Jeremiah M. Faries  
Department of Psychology  
Northwestern University  
Evanston, IL 60208

Dr. Beatrice J. Farr  
Army Research Institute  
PERI-IC  
5001 Eisenhower Avenue  
Alexandria, VA 22333

Dr. Marshall J. Farr, Consultant  
Cognitive & Instructional Sciences  
2520 North Vernon Street  
Arlington, VA 22207

Dr. P.A. Federico  
Code 51  
NPRDC  
San Diego, CA 92152-6800

Dr. Paul Feltoich  
Southern Illinois University  
School of Medicine  
Medical Education Department  
P.O. Box 3926  
Springfield, IL 62708

Dr. Elizabeth Fennema  
Curriculum and Instruction  
University of Wisconsin  
225 North Mills Street  
Madison, WI 53706

Dr. Michael Flanigan  
Code 52  
NPRDC  
San Diego, CA 92152-6800

Dr. J. D. Fletcher  
Institute for Defense Analyses  
1801 N. Beauregard St.  
Alexandria, VA 22311

Dr. Carl H. Frederiksen  
Dept. of Educational Psychology  
McGill University  
3700 McTavish Street  
Montreal, Quebec  
CANADA H3A 1Y2

Dr. John R. Frederiksen  
c/o Dr. Barbara White  
School of Education  
Tolman Hall, EMST  
University of California  
Berkeley, CA 94720

Department of Humanities and  
Social Sciences  
Harvey Mudd College  
Claremont, CA 91711

Dr. Alfred R. Frey  
AFOSR/NL, Bldg. 410  
Bolling AFB, DC 20332-6448

Dr. Michael Friendly  
Psychology Department  
York University  
Toronto ONT  
CANADA M3J 1P3

Col. Dr. Ernst Fries  
Heerespsychologischer Dienst  
Maria Theresien-Kaserne  
1130 Wien  
AUSTRIA

Dr. Donald R. Gentner  
Philips Laboratories  
345 Scarborough Road  
Briarcliff Manor, NY 10510

Dr. Alan S. Gevins  
EEG Systems Laboratory  
51 Federal Street, Suite 401  
San Francisco, CA 94107

Mr. Lee Gladwin  
305 Davis Avenue  
Leesburg, VA 22075

Dr. Arthur M. Glenberg  
University of Wisconsin  
W. J. Brogden Psychology Bldg.  
1202 W. Johnson Street  
Madison, WI 53706

Dr. Joseph Goguen  
Computer Science Laboratory  
SRI International  
333 Ravenswood Avenue  
Menlo Park, CA 94025

Dr. Paul E. Gold  
University of Virginia  
Department of Psychology  
Charlottesville, VA 22903

Dr. Patricia Goldman-Rakic  
Section of Neuroanatomy, C303  
Yale University  
333 Cedar Street  
New Haven, CT 06510

Dr. Timothy Goldsmith  
Department of Psychology  
University of New Mexico  
Albuquerque, NM 87131

Dr. Sherrie Gott  
AFHRL/MOMJ  
Brooks AFB, TX 78235-5401

Dr. Richard H. Granger  
Department of Computer Science  
University of California, Irvine  
Irvine, CA 92717

Prof. Edward Haertel  
School of Education  
Stanford University  
Stanford, CA 94305

Dr. Bruce W. Hamill  
Research Center  
The Johns Hopkins University  
Applied Physics Laboratory  
Johns Hopkins Road  
Laurel, MD 20707

Dr. Patrick R. Harrison  
Computer Science Department  
U.S. Naval Academy  
Annapolis, MD 21402-5002

Janice Hart  
Office of the Chief  
of Naval Operations  
OP-111J2  
Department of the Navy  
Washington, D.C. 20350-2000

Dr. Frederick Hayes-Roth  
Teknowledge  
P.O. Box 10119  
1850 Embarcadero Rd.  
Palo Alto, CA 94303

Ms. Julia S. Hough  
Cambridge University Press  
40 West 20th Street  
New York, NY 10011

Dr. William Howell  
Chief Scientist  
AFHRL/CA  
Brooks AFB, TX 78235-5601

Dr. Eva Hudlicka  
BBN Laboratories  
10 Moulton Street  
Cambridge, MA 02238

Dr. Steven Hunka  
3-104 Educ. N.  
University of Alberta  
Edmonton, Alberta  
CANADA T6G 2G5

Dr. Jack Hunter  
2122 Coolidge Street  
Lansing, MI 48906

Dr. Ed Hutchins  
Intelligent Systems Group  
Institute for  
Cognitive Science (C-015)  
UCSD  
La Jolla, CA 92093

Dr. Martin J. Ippel  
Postbus 9555  
2300 RB Leiden  
THE NETHERLANDS

Dr. Alice M. Isen  
Department of Psychology  
University of Maryland  
Collegeville, MD 21228

Dr. Robin Jeffries  
Hewlett-Packard Laboratories, 3L  
P.O. Box 10490  
Palo Alto, CA 94303-0971

Dr. Peder Johnson  
Department of Psychology  
University of New Mexico  
Albuquerque, NM 87131

Dr. Michael Kaplan  
Office of Basic Research  
U.S. Army Research Institute  
5001 Eisenhower Avenue  
Alexandria, VA 22333-5600

Dr. Demetrios Karis  
GTE Labs, MS 61  
40 Sylvan Road  
Waltham, MA 02254

Dr. J.A.S. Kello  
Center for Complex Systems  
Building MT 9  
Florida Atlantic University  
Boca Raton, FL 33431

Dr. David Kieras  
Technical Communication Program  
TIDAL Bldg., 2360 Bonisteel Blvd.  
University of Michigan  
Ann Arbor, MI 48109-2108

Dr. Walter Kintoch  
Department of Psychology  
University of Colorado  
Boulder, CO 80309-0345

Dr. Richard J. Koubek  
Department of Biomedical  
& Human Factors  
139 Engineering & Math Bldg.  
Wright State University  
Dayton, OH 45435

Dr. Yuh-Jeng Lee  
Department of Computer Science  
Code 52Le  
Naval Postgraduate School  
Monterey, CA 93943

Dr. Alan M. Leagold  
Learning R&D Center  
University of Pittsburgh  
Pittsburgh, PA 15260

Dr. Michael Levine  
Educational Psychology  
210 Education Bldg.  
University of Illinois  
Champaign, IL 61801

Dr. Jack Lochhead  
University of  
Massachusetts  
Physics Department  
Amherst, MA 01003

Dr. Jane Malin  
Mail Code EF5  
NASA Johnson Space Center  
Houston, TX 77058

Dr. Sandra P. Marshall  
Dept. of Psychology  
San Diego State University  
San Diego, CA 92182

Dr. Manton M. Matthews  
Department of Computer Science  
University of South Carolina  
Columbia, SC 29208

Dr. Michael McNeese  
DET-1, ALUMED  
BLJG 248  
Wright-Patterson AFB, OH 45432

Dr. Barbara Means  
SRI International  
333 Ravenswood Avenue  
Menlo Park, CA 94025

Dr. Douglas L. Medin  
Department of Psychology  
University of Michigan  
Ann Arbor, MI 48109

Prof. Jacques Meblar  
5-1, bd. Raspail  
75006 Paris  
FRANCE

Dr. Arthur Meisner  
Computer Arts and  
Education Laboratory  
New York University  
719 Broadway, 12th floor  
New York, NY 10003

Dr. Jose Mestre  
Department of Physics  
Hasbrouck Laboratory  
University of Massachusetts  
Amherst, MA 01003

Dr. Joel A. Michael  
Department of Physiology  
Rush Presbyterian-St. Luke's  
Medical Center  
Rush Medical College  
Chicago, IL 60612

Dr. Jason Millman  
Department of Education  
Roberts Hall  
Cornell University  
Ithaca, NY 14853

Dr. Robert Mislevy  
Educational Testing Service  
Princeton, NJ 08541

Dr. William Montague  
NPRDC Code 13  
San Diego, CA 92152-6800

Dr. Melvin D. Montemerlo  
NASA Headquarters  
Code RC  
Washington, DC 20546

Dr. Johanna D. Moore  
LRDC  
University of Pittsburgh  
3939 O'Hara Street  
Pittsburgh, PA 15260

Dr. Randy Mumaw  
Human Sciences  
Westinghouse Science  
& Technology Ctr.  
1310 Beulah Road  
Pittsburgh, PA 15235

Mr. J. Nelissen  
Twente University of Technology  
Fac. Biol. Toespeste Onderwyskurde  
P. O. Box 217  
7500 AE Enschede  
The NETHERLANDS

Dr. T. Niblett  
The Turing Institute  
George House  
36 North Hanover Street  
Glasgow G1 2AD  
UNITED KINGDOM

Dr. A. F. Norcio  
Code 5530  
Naval Research Laboratory  
Washington, DC 20375-5000

Dr. Donald A. Norman  
C-015  
Institute for Cognitive Science  
University of California  
La Jolla, CA 92093

Library, NPRDC  
Code P201L  
San Diego, CA 92152-6800

Librarian  
Naval Center for Applied Research  
in Artificial Intelligence  
Naval Research Laboratory  
Code 5510  
Washington, DC 20375-5000

Dr. Judith Reitman Olson  
Graduate School of Business  
University of Michigan  
Ann Arbor, MI 48109-1234

Office of Naval Research,  
Code 1142CS  
800 N. Quincy Street  
Arlington, VA 22217-5000  
(6 Copies)

Dr. Judith Orasanu  
Basic Research Office  
Army Research Institute  
5001 Eisenhower Avenue  
Alexandria, VA 22333

Dr. Ray S. Perez  
ARI (PERI-II)  
5001 Eisenhower Avenue  
Alexandria, VA 22333

Dr. Nancy N. Perry  
Naval Education and Training  
Program Support Activity  
Code-047  
Building 2435  
Pensacola, FL 32509-5000

Dept. of Administrative Sciences  
Code 54  
Naval Postgraduate School  
Monterey, CA 93943-5026

Dr. Lynne Roder  
Department of Psychology  
Carnegie-Mellon University  
Schenley Park  
Pittsburgh, PA 15213

Dr. Stephen Roder  
NWREL  
101 SW Main, Suite 500  
Portland, OR 97204

Dr. Charles M. Reigekuth  
330 Huntington Hall  
Syracuse University  
Syracuse, NY 13244

Dr. Lauren Resnick  
Learning R. & D. Center  
University of Pittsburgh  
3939 O'Hara Street  
Pittsburgh, PA 15213

Dr. Edwin L. Riesland  
Dept. of Computer and  
Information Science  
University of Massachusetts  
Amherst, MA 01003

Dr. W. A. Rizzo  
Head, Human Factors Division  
Naval Training Systems Center  
Code 26  
12350 Research Parkway  
Orlando, FL 32826-3224

Mr. William A. Rizzo  
Code 71  
Naval Training Systems Center  
Orlando, FL 32813

LT CDR Michael N. Rodgers  
Canadian Forces Personnel  
Applied Research Unit  
4900 Yonge Street, Suite 600  
Willowdale, Ontario M2N 6B7  
CANADA

Dr. Fumiko Semejima  
Department of Psychology  
University of Tennessee  
310B Austin Peay Bldg.  
Knoxville, TN 37916-0900

Dr. James F. Sanford  
Department of Psychology  
George Mason University  
4400 University Drive  
Fairfax, VA 22030

Lowell Schoer  
Psychological & Quantitative  
Foundations  
College of Education  
University of Iowa  
Iowa City, IA 52242

Dr. Judith W. Segal  
OERI  
555 New Jersey Ave., NW  
Washington, DC 20208

Dr. Robert J. Seidel  
US Army Research Institute  
5001 Eisenhower Ave.  
Alexandria, VA 22333

Dr. Michael G. Shafro  
NASA Ames Research Ctr.  
Mail Stop 239-1  
Moffett Field, CA 94035

Dr. Valerie L. Shalin  
Department of Industrial  
Engineering  
State University of New York  
342 Lawrence D. Bell Hall  
Buffalo, NY 14260

Mr. Colin Sheppard  
AXC2 Block 3  
Admiralty Research Establishment  
Ministry of Defence Portdown  
Portsmouth Harb P064AA  
UNITED KINGDOM

Dr. Ben Shneiderman  
Dept. of Computer Science  
University of Maryland  
College Park, MD 20742

Dr. Randall Shumaker  
Naval Research Laboratory  
Code 5510  
4555 Overlook Avenue, S.W.  
Washington, DC 20375-5000

Dr. Edward E. Smith  
Department of Psychology  
University of Michigan  
330 Packard Road  
Ann Arbor, MI 48103

Dr. Alfred F. Smode  
Code 7A  
Research and Development Dept.  
Naval Training Systems Center  
Orlando, FL 32813-7100

Dr. Larry Squire  
Department of Psychiatry  
Veterans Medical Research Foundation  
University of California  
3350 La Jolla Village Dr.  
San Diego, CA 92161

N. S. Sridharan  
FMC Corporation  
Box 580  
1205 Coleman Avenue  
Santa Clara, CA 95052

Dr. James J. Staszewski  
Dept. of Psychology  
University of South Carolina  
Columbia, SC 29210

Dr. Marian Stearns  
SRJ International  
333 Ravenswood Ave.  
Room B-5124  
Menlo Park, CA 94025

Dr. Ted Steinke  
Dept. of Geography  
University of South Carolina  
Columbia, SC 29208

Dr. David E. Stone  
Computer Teaching Corporation  
1713 South Neil Street  
Urbana, IL 61820

Mr. Michael J. Strait  
UMUC Graduate School  
College Park, MD 20742

Dr. Sharon Tkacz  
Allen Corporation  
209 Madison Street  
Alexandria, VA 22314

Dr. Harold P. Van Cott  
Committee on Human Factors  
National Academy of Sciences  
2101 Constitution Avenue  
Washington, DC 20418

Dr. Frank L. Vicino  
Navy Personnel R&D Center  
San Diego, CA 92152-6800

Dr. Ross R. Vickers  
Stress Medicine Department  
P.O. Box 85122  
Naval Health Research Center  
San Diego, CA 92186-5122

Dr. Jerry Vogt  
Department of Psychology  
St. Norbert College  
De Pere, WI 54115-2099

Dr. Thomas A. Warm  
FAA Academy AAC934D  
P.O. Box 25082  
Oklahoma City, OK 73125

Dr. Beth Warren  
BBN Laboratories, Inc.  
10 Moulton Street  
Cambridge, MA 02238

Dr. Diane Wearne  
Department of Educational  
Development  
University of Delaware  
Newark, DE 19711

Dr. Douglas Wetzel  
Code 51  
Navy Personnel R&D Center  
San Diego, CA 92152-6800

Dr. Barbara White  
School of Education  
Tolman Hall, EMST  
University of California  
Berkeley, CA 94720

Dr. Mark Wilson  
School of Education  
University of California  
Berkeley, CA 94720

Dr. Eugene Winograd  
Department of Psychology  
Emory University  
Atlanta, GA 30322

Dr. Robert A. Winber  
U.S. Army Institute for the  
Behavioral and Social Sciences  
5001 Eisenhower Avenue  
Alexandria, VA 22333-5600

Dr. Merita C. Witzrock  
Graduate School of Education  
UCLA  
Los Angeles, CA 90024

Dr. Wallace Wulfelt, III  
Navy Personnel R&D Center  
Code 51  
San Diego, CA 92152-6800

Dr. Joseph L. Young  
National Science Foundation  
Room 320  
1800 G Street, N.W.  
Washington, DC 20550